

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 13:11:42 : Search time 55 Seconds  
(without alignments)  
4510.936 Million cell updates/sec

Title: US-09-822-080B-1  
Perfect score: 809  
Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	4.0	623	US-09-385-982-207	Sequence 207, App
2	31	3.8	1693	5169835-3	Patent No. 5169835
3	31	3.8	1936	US-08-836-567-5	Sequence 5, Appl
4	31	3.8	2793	US-08-836-567-7	Sequence 7, Appl
5	30	3.7	39	US-09-619-103-4	Sequence 4, Appl
6	30	3.7	361	US-08-465-380-39	Sequence 39, Appl
7	30	3.7	361	US-08-486-397-39	Sequence 39, Appl
8	30	3.7	361	US-08-486-399-39	Sequence 39, Appl
9	30	3.7	361	US-08-461-965-39	Sequence 39, Appl
10	30	3.7	361	US-08-634-641-39	Sequence 39, Appl
11	30	3.7	361	US-09-249-471-39	Sequence 39, Appl
12	30	3.7	361	US-09-249-472-39	Sequence 39, Appl
13	30	3.7	361	US-09-249-451-39	Sequence 39, Appl
14	30	3.7	361	US-08-809-455-39	Sequence 39, Appl
15	30	3.7	361	US-09-249-461-39	Sequence 39, Appl
16	30	3.7	361	US-09-249-448-39	Sequence 39, Appl
17	30	3.7	770	US-09-385-982-542	Sequence 542, App
18	30	3.7	1505	US-08-909-965C-13	Sequence 13, Appl
19	30	3.7	1651	US-09-465-558-49	Sequence 49, Appl
20	30	3.7	2274	US-09-388-743-17	Sequence 17, Appl
21	30	3.7	2367	US-08-441-139-3	Sequence 3, Appl
22	30	3.7	2658	US-08-592-383-3	Sequence 3, Appl
23	30	3.7	2940	US-08-592-383-1	Sequence 1, Appl
24	30	3.7	2940	5171671-1	Patent No. 5171671
25	30	3.7	3227	US-08-372-892-3	Sequence 3, Appl
26	30	3.7	3254	US-08-372-892-1	Sequence 1, Appl
27	30	3.7	4665	US-08-948-378A-7	Sequence 7, Appl

28	30	3.7	4665	4	US-09-169-425C-7	Sequence 7, Appl
29	29	3.6	104	2	US-08-803-899-8	Sequence 8, Appl
30	29	3.6	255	4	US-09-480-921B-26	Sequence 26, Appl
31	29	3.6	779	1	US-07-841-846-22	Sequence 22, Appl
32	29	3.6	779	1	US-08-147-023-22	Sequence 22, Appl
33	29	3.6	779	1	US-08-449-570-22	Sequence 22, Appl
34	29	3.6	779	2	US-08-449-700-22	Sequence 22, Appl
35	29	3.6	779	2	US-08-449-699A-22	Sequence 22, Appl
36	29	3.6	1106	3	PCT-US96-362-318-1	Sequence 1, Appl
37	29	3.6	1166	5	PCT-US96-12129B-1	Sequence 1, Appl
38	29	3.6	1283	4	US-09-282-305-11	Sequence 11, Appl
39	29	3.6	1374	1	US-08-123-761A-2	Sequence 2, Appl
40	29	3.6	1400	4	US-09-245-281-40	Sequence 40, Appl
41	29	3.6	1400	4	US-09-207-359B-40	Sequence 40, Appl
42	29	3.6	1485	4	US-09-372-422A-39	Sequence 39, Appl
43	29	3.6	1502	2	US-08-651-940-1	Sequence 1, Appl
44	29	3.6	1502	4	US-09-295-029-1	Sequence 1, Appl
45	29	3.6	1664	1	US-07-863-169A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-09-385-982-207/c  
; Sequence 207, Application US/09385982  
; Patent No. 6262334  
; GENERAL INFORMATION:  
; APPLICANT: ENDEGE, WILSON O., ET AL.  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; TITLE OF INVENTION: PRODUCTS: II  
; FILE REFERENCE: CDNA-260XX  
; CURRENT APPLICATION NUMBER: US/09/385,982  
; CURRENT FILING DATE: 1999-08-30  
; EARLIER APPLICATION NUMBER: 09/328,111  
; EARLIER FILING DATE: 1999-06-08  
; EARLIER APPLICATION NUMBER: 60/117,393  
; EARLIER FILING DATE: 1999-01-27  
; EARLIER APPLICATION NUMBER: 60/098,639  
; EARLIER FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 544  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 207  
; LENGTH: 623  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(623)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-385-982-207

Query Match 4.0%; Score 32; DB 4; Length 623;  
Best Local Similarity 100.0%; Pred. No. 2.2e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 778 GTTCTTAAAAA.....AAAAAAAAAAAAA 809  
Db 38 GTTCTTAAAAA.....AAAAAAAAAAAAA 7

RESULT 2  
5169835-3  
; Patent No. 5169835  
; APPLICANT: WAI-YEE, CHAN  
; TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS  
; NUMBER OF SEQUENCES: 48  
; CURRENT APPLICATION DATA:  
; FILING DATE: 07-AUG-1989  
; SEQ ID NO:3  
; LENGTH: 1693  
5169835-3

Query Match 3.8%; Score 31; DB 3; Length 1926;  
Best Local Similarity 100.0%; Pred. No. 5.6e-05;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTTAAAAA 809  
DB 1894 TTCTTAAAAA 1924

RESULT 4  
US-08-836-567-7  
; Sequence 7, Application US/08836567  
; Patent No. 6130367  
; GENERAL INFORMATION:  
; APPLICANT: Kossmann, Jens  
; APPLICANT: Springer, Franziska  
; APPLICANT: Abel, Gernot  
; TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES  
; TITLE OF INVENTION: INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC  
; TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836.567  
; FILING DATE: 24-JUL-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP95/04415  
; FILING DATE: 09-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 44 41 408.0  
; FILING DATE: 10-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: Agrevo-4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2793 base pairs  
; TYPE: nucleotide  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ORIGINAL SOURCE:  
; ORGANISM: Solanum tuberosum  
; STRAIN: cv Dsire  
; TISSUE TYPE: leaf tissue  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA-library in Lambda ZAPII  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 242..2542  
; US-08-836-567-7

Query Match 3.8%; Score 31; DB 3; Length 2793;  
Best Local Similarity 100.0%; Pred. No. 5.4e-05;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 3.8%; Score 31; DB 6; Length 1693;  
Best Local Similarity 100.0%; Pred. No. 5.7e-05;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTTAAAAA 809  
DB 1660 TTCTTAAAAA 1690

RESULT 3  
US-08-836-567-5  
; Sequence 5, Application US/08836567  
; Patent No. 6130367  
; GENERAL INFORMATION:  
; APPLICANT: Kossmann, Jens  
; APPLICANT: Springer, Franziska  
; APPLICANT: Abel, Gernot  
; TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES  
; TITLE OF INVENTION: INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC  
; TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10020  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836.567  
; FILING DATE: 24-JUL-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP95/04415  
; FILING DATE: 09-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 44 41 408.0  
; FILING DATE: 10-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haley Jr., James F.  
; REGISTRATION NUMBER: 27,794  
; REFERENCE/DOCKET NUMBER: Agrevo-4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1926 base pairs  
; TYPE: nucleotide  
; STRANDEDNESS: unknown  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Solanum tuberosum  
; STRAIN: cv. Berolina  
; TISSUE TYPE: tuber tissue  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA-library in pBluescriptSk+  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2..1675  
; OTHER INFORMATION: /function= "Polymerization of  
; OTHER INFORMATION: starch"  
; OTHER INFORMATION: /product= "Starch synthase"  
; US-08-836-567-5

QY 779 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809  
 Db 2761 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2791

RESULT 5  
 US-09-619-103-4  
 ; Sequence 4, Application US/09619103  
 ; Patent No. 6429300  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kurz, Markus  
 ; APPLICANT: Lohse, Peter  
 ; APPLICANT: Wagner, Richard  
 ; TITLE OF INVENTION: Peptide Acceptor Ligation Methods  
 ; FILE REFERENCE: 50036/031002  
 ; CURRENT APPLICATION NUMBER: US/09/619,103  
 ; CURRENT FILING DATE: 2000-07-19  
 ; PRIOR APPLICATION NUMBER: 60/145,834  
 ; PRIOR FILING DATE: 1999-07-27  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 39  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: designed sequence to act as a linker  
 US-09-619-103-4

Query Match 3.7% Score 30; DB 4; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 0.00023;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809  
 Db 7 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 36

RESULT 6  
 US-08-465-380-39  
 ; Sequence 39, Application US/08465380  
 ; Patent No. 5863894  
 ; GENERAL INFORMATION:  
 ; APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
 ; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
 ; APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
 ; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,  
 ; APPLICANT: Peter W. Bergum  
 ; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
 ; TITLE OF INVENTION: PROTEIN  
 ; NUMBER OF SEQUENCES: 356  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; STREET: Suite 4700  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/465,380  
 ; FILING DATE: June 5, 1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/326,110  
 ; FILING DATE: October 18, 1994  
 ; ATTORNEY/AGENT INFORMATION:

; NAME: BIGGS, SUZANNE L.  
 ; REGISTRATION NUMBER: 30,158  
 ; REFERENCE/DOCKET NUMBER: 213/268  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 361 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Necator americanus  
 ; FEATURE:  
 ; NAME/KEY: Coding Sequence  
 ; LOCATION: 16....252  
 US-08-465-380-39

Query Match 3.7% Score 30; DB 2; Length 361;  
 Best Local Similarity 100.0%; Pred. No. 0.00018;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809  
 Db 325 TTCTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354

RESULT 7  
 US-08-486-397-39  
 ; Sequence 39, Application US/08486397  
 ; Patent No. 5866542  
 ; GENERAL INFORMATION:  
 ; APPLICANT: George P. Vlasuk, Patric H. Stanssens,  
 ; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,  
 ; APPLICANT: Yves R. Laroche, Laurent S. Jespers,  
 ; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,  
 ; APPLICANT: Peter W. Bergum  
 ; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT  
 ; TITLE OF INVENTION: PROTEIN  
 ; NUMBER OF SEQUENCES: 357  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: 633 West Fifth Street  
 ; STREET: Suite 4700  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/486,397  
 ; FILING DATE: June 5, 1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/326,110  
 ; FILING DATE: October 18, 1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BIGGS, SUZANNE L.  
 ; REGISTRATION NUMBER: 30,158  
 ; REFERENCE/DOCKET NUMBER: 213/269  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 39:  
 ; SEQUENCE CHARACTERISTICS:

Mon Jan 27 12:25:13 2003

```

; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
;
US-08-486-397-39
    Query Match          3.7%; Score 30; DB 2; Length 361;
    Best Local Similarity 100.0%; Pred. No. 0.00018;
    Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354

RESULT 8
US-08-486-399-39
; Sequence 39, Application US/08486399
; Patent No. 5866543
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Berquum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,399
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 213/270
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
;
US-08-461-965-39
    Query Match          3.7%; Score 30; DB 2; Length 361;
    Best Local Similarity 100.0%; Pred. No. 0.00018;
    Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354

RESULT 9
US-08-461-965-39
; Sequence 39, Application US/08461965
; Patent No. 5872098
; GENERAL INFORMATION:
; APPLICANT: George P. Vlasuk, Patric H. Stanssens,
; APPLICANT: Joris H.L. Mensens, Marc J. Lauwereys,
; APPLICANT: Yves R. Laroche, Laurent S. Jespers,
; APPLICANT: Yannick G.J. Gansemans, Matthew Moyle,
; APPLICANT: Peter W. Berquum
; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,965
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 210/243
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
;
US-08-461-965-39
    Query Match          3.7%; Score 30; DB 2; Length 361;
    Best Local Similarity 100.0%; Pred. No. 0.00018;
    Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354
```



## RESULT 10

US-08-634-641-39  
; Sequence 39, Application US/08634641  
; Patent No. 5955294

## GENERAL INFORMATION:

; APPLICANT: Vlasuk, George P. Vlasuk  
; APPLICANT: Stanssens, Patrick Eric Hugo  
; APPLICANT: Mensens, Joris Hilda Lieven  
; APPLICANT: Lauwereys, Marc Josef  
; APPLICANT: Laroche, Yves Rene  
; APPLICANT: Jespers, Laurent Stephane  
; APPLICANT: Ganssemans, Yannick Georges Jozef  
; APPLICANT: Moyle, Matthew  
; APPLICANT: Bergum, Peter W.

; TITLE OF INVENTION: NEMATODE-EXTRACTED ANTICOAGULANT

; TITLE OF INVENTION: PROTEIN

; NUMBER OF SEQUENCES: 356

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; STREET: Suite 4700

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/634,641

; FILING DATE: April 19, 1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/13231

; FILING DATE: October 17, 1995

; APPLICATION NUMBER: 08/486,399

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/486,397

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/465,380

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/461,965

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/326,110

; FILING DATE: October 18, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BIGGS, SUZANNE L.

; REGISTRATION NUMBER: 30,158

; REFERENCE/DOCKET NUMBER: 219/136

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 39:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; ORGANISM: Necator americanus

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 16...252

; US-08-634-641-39

Query Match 3.7%; Score 30; DB 2; Length 361;

Best Local Similarity 100.0%; Pred. No. 0.00018;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 809

|||||

Db 325 TTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 354

## RESULT 11

US-09-249-471-39

; Sequence 39, Application US/09249471

; Patent No. 6040441

; GENERAL INFORMATION:

; APPLICANT: Vlasuk, George Phillip

; APPLICANT: Stanssens, Patrick Eric Hugo

; APPLICANT: Messens, Joris Hilda Lieven

; APPLICANT: Lauwereys, Marc Josef

; APPLICANT: Laroche, Yves Rene

; APPLICANT: Jespers, Laurent Stephane

; APPLICANT: Ganssemans, Yannick Georges Jozef

; APPLICANT: Moyle, Matthew

; APPLICANT: Bergum, Peter W.

; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE

; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT

; NUMBER OF SEQUENCES: 356

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; STREET: Suite 4700

; CITY: Los Angeles

; STATE: California

; COUNTRY: U.S.A.

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/249,471

; FILING DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/809,455

; FILING DATE: April 17, 1997

; APPLICATION NUMBER: PCT/US95/13231

; FILING DATE: October 17, 1995

; APPLICATION NUMBER: 08/486,399

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/486,397

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/465,380

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/461,965

; FILING DATE: June 5, 1995

; APPLICATION NUMBER: 08/326,110

; FILING DATE: October 18, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: BIGGS, SUZANNE L.

; REGISTRATION NUMBER: 30,158

; REFERENCE/DOCKET NUMBER: 216/270

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 39:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; ORIGINAL SOURCE:

; ORGANISM: Necator americanus

; FEATURE:

```
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
US-09-249-471-39

Query Match          3.7%; Score 30; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAATAAAAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAATAAAAAAAAAAAAAAAAAAAAAA 354

RESULT 12
US-09-249-472-39
; Sequence 39, Application US/09249472
; Patent No. 6046318
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansmans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,472
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
US-09-249-472-39

Query Match          3.7%; Score 30; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAATAAAAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAATAAAAAAAAAAAAAAAAAAAAAA 354

RESULT 13
US-09-249-451-39
; Sequence 39, Application US/09249451
; Patent No. 6087487
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansmans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,451
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
```

```
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
; US-09-249-451-39

Query Match 3.7%; Score 30; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 354

RESULT 14
US-08-809-455-39
; Sequence 39, Application US/08809455
; Patent No. 6090916
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE:
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,455
; FILING DATE: April 17, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/486,397
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/465,380
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/461,965
; FILING DATE: June 5, 1995
```

```
; APPLICATION NUMBER: 08/326,110
; FILING DATE: October 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BIGGS, SUZANNE L.
; REGISTRATION NUMBER: 30,158
; REFERENCE/DOCKET NUMBER: 216/270
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; ORGANISM: Necator americanus
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 16...252
; US-08-809-455-39

Query Match 3.7%; Score 30; DB 3; Length 361;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 TTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 325 TTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 354

RESULT 15
US-09-249-461-39
; Sequence 39, Application US/09249461
; Patent No. 6096877
; GENERAL INFORMATION:
; APPLICANT: Vlasuk, George Phillip
; APPLICANT: Stanssens, Patrick Eric Hugo
; APPLICANT: Messens, Joris Hilda Lieven
; APPLICANT: Lauwereys, Marc Josef
; APPLICANT: Laroche, Yves Rene
; APPLICANT: Jespers, Laurent Stephane
; APPLICANT: Gansemans, Yannick Georges Jozef
; APPLICANT: Moyle, Matthew
; APPLICANT: Bergum, Peter W.
; TITLE OF INVENTION: NEMATODE-EXTRACTED SERINE PROTEASE
; TITLE OF INVENTION: INHIBITORS AND ANTICOAGULANT
; NUMBER OF SEQUENCES: 356
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,461
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/809,455
; FILING DATE: April 17, 1997
; APPLICATION NUMBER: PCT/US95/13231
; FILING DATE: October 17, 1995
; APPLICATION NUMBER: 08/486,399
```

;; FILING DATE: June 5, 1995  
;; APPLICATION NUMBER: 08/486,397  
;; FILING DATE: June 5, 1995  
;; APPLICATION NUMBER: 08/465,380  
;; FILING DATE: June 5, 1995  
;; APPLICATION NUMBER: 08/461,965  
;; FILING DATE: June 5, 1995  
;; APPLICATION NUMBER: 08/326,110  
;; FILING DATE: October 18, 1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BIGGS, SUZANNE L.  
;; REGISTRATION NUMBER: 30,158  
;; REFERENCE/DOCKET NUMBER: 216/270  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (213) 489-1600  
;; TELEFAX: (213) 955-0440  
;; TELEX: 67-3510  
;; INFORMATION FOR SEQ ID NO: 39:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 361 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; ORIGINAL SOURCE:  
;; ORGANISM: Necator americanus  
;; FEATURE:  
;; NAME/KEY: Coding Sequence  
;; LOCATION: 16....252  
US-09-249-461-39

Query Match 3.7% Score 30; DB 3; Length 361;  
Best Local Similarity 100.0%; Pred. No. 0.00018;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 780 TTCTAAAAAAAAAAAAAAAAAAAAAAAA 809  
|||  
Db 325 TTCTAAAAAAAAAAAAAAAAAAAAAAAA 354

Search completed: January 25, 2003, 15:49:53  
Job time : 65 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 15:49:58 : Search time 2183 Seconds  
(without alignments)  
6001.900 Million cell updates/sec

Title: US-09-822-080B-1

Perfect score: 809

Sequence: 1 gatgaatccatgtttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	69	8.5	685	17	BH711815
C 2	69	8.5	830	17	BH606894
C 3	38	4.7	660	17	BH456221
C 4	36	4.4	475	14	BQ157439
C 5	36	4.4	666	17	BH678808
C 6	35	4.3	240	10	BE209488

C	7	35	4.3	284	9	AI826026
	8	35	4.3	292	10	AW432208
	9	35	4.3	325	13	BI344610
C	10	35	4.3	342	9	AI268815
C	11	35	4.3	439	9	AI957804
C	12	35	4.3	542	9	AJ500337
C	13	35	4.3	575	9	AL629881
C	14	34	4.2	135	9	AA584642
	15	34	4.2	202	10	AW101025
	16	34	4.2	234	13	BI534973
C	17	34	4.2	240	9	AI567012
	18	34	4.2	245	9	AU165623
	19	34	4.2	277	14	D24598
	20	34	4.2	288	9	AL047763
C	21	34	4.2	347	12	BG628219
	22	34	4.2	362	10	BE617172
	23	34	4.2	375	9	AU071066
C	24	34	4.2	427	9	AI921506
C	25	34	4.2	467	9	AU070451
C	26	34	4.2	478	14	BQ800172
	27	34	4.2	574	13	BI889646
C	28	34	4.2	601	14	BQ409166
	29	34	4.2	656	12	BF531023
	30	34	4.2	751	12	BF141790
	31	34	4.2	759	13	BI520175
	32	33	4.1	167	10	AW459222
C	33	33	4.1	198	9	AI463029
C	34	33	4.1	206	9	AI686844
C	35	33	4.1	210	10	AW656286
	36	33	4.1	267	13	BI837299
C	37	33	4.1	348	9	AA485710
C	38	33	4.1	372	9	AI593773
C	39	33	4.1	373	12	BG627869
	40	33	4.1	377	12	BE849053
C	41	33	4.1	387	13	BM072677
	42	33	4.1	399	14	BQ087397
	43	33	4.1	422	13	BM255161
C	44	33	4.1	467	13	BI865527
C	45	33	4.1	484	14	BQ479833

#### ALIGNMENTS

RESULT 1	BH711815	685 bp	DNA	linear	GSS 20-FEB-2002			
LOCUS	BOHYO04TR	BO_2_3_KB	Brassica oleracea	genomic clone	BOHYO04, DNA			
DEFINITION	sequence.							
ACCESSION	BH711815							
VERSION	BH711815.1	GI:18802293						
KEYWORDS	GSS.							
SOURCE	Brassica oleracea.							
ORGANISM	Brassica oleracea							
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.							
AUTHORS	Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.							
TITLE	Whole genome shotgun sequencing of Brassica oleracea							
JOURNAL	Unpublished (2001)							
COMMENT	Other_GSSs: BOHYO04TF Contact: Chris Town							
FEATURES	TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA. Tel: 301-838-3523 Fax: 301-838-0208 Email: cdtown@tigr.org DNA is from a doubled haploid provided by Tom Osborn. Seq primer: TR Class: sheared ends. Location/Qualifiers 1..685							

```

/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHY004"
/clone_lib="BO_2_3_KB"
/notes="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      176 a  164 c  171 g  174 t
ORIGIN

Query Match
Best Local Similarity 8.5%; Score 69; DB 17; Length 685;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 TGGTCTTACGACAGAGCTGGAGGAGAGATTCTGCGAGGAGTTTCTTCCAGATCAAGAA 447
|||||
DB 381 TGGTCTTACGACAGAGCTGGAGGAGAGATTCTGCGAGGAGTTTCTTCCAGATCAAGAA 440
|||||
QY 448 AACTGAAGC 456
|||||
DB 441 AACTGAAGC 449

RESULT 2
BH606894/c
LOCUS
DEFINITION      BH606894      830 bp      DNA      linear      GSS 15-DEC-2001
BOGFM83TR BOGF Brassica oleracea genomic clone BOGFM83, DNA
sequence.
ACCESSION      BH606894
VERSION
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 830)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOGFM83TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..830
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGFM83"
/clone_lib="BOGF"
/notes="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      213 a  185 c  193 g  239 t
ORIGIN

Query Match
Best Local Similarity 8.5%; Score 69; DB 17; Length 830;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 TGGTCTTACGACAGAGCTGGAGGAGAGATTCTGCGAGGAGTTTCTTCCAGATCAAGAA 447
|||||
DB 234 TGGTCTTACGACAGAGCTGGAGGAGAGATTCTGCGAGGAGTTTCTTCCAGATCAAGAA 175
|||||
QY 448 AACTGAAGC 456
|||||
DB 174 AACTGAAGC 166
|||||

```

```

RESULT 3
BH456221/c
LOCUS
DEFINITION      BH456221      660 bp      DNA      linear      GSS 12-DEC-2001
BOHRQ52TF BOHR Brassica oleracea genomic clone BOHRQ52, DNA
sequence.
ACCESSION      BH456221
VERSION
KEYWORDS
SOURCE
ORGANISM
Brassica oleracea.
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 660)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOHRQ52TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..660
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHRQ52"
/clone_lib="BOHR"
/notes="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      193 a  163 c  145 g  159 t
ORIGIN

Query Match
Best Local Similarity 4.7%; Score 38; DB 17; Length 660;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 623 TCCAAGACTATGCTATTATCTCGAGAGAAATTAAAGAC 660
|||||
DB 125 TCCAAGACTATGCTATTATCTCGAGAGAAATTAAAGAC 88
|||||

RESULT 4
BH157439
LOCUS
DEFINITION      BH157439      475 bp      mRNA      linear      EST 24-APR-2002
NF104H1IRIF1096 Irradiated Medicago truncatula cDNA clone
NF104H1IR 5', mRNA sequence.
ACCESSION      BH157439
VERSION
KEYWORDS
SOURCE
ORGANISM
barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 475)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula irradiated library
Unpublished (2001)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation

```

2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7391  
 Fax: 580 221 7380  
 Email: gdmay@noble.org  
 Insert Length: 475 Std Error: 0.00  
 Plate: 104 row: H column: 11  
 Seq primer: TCACACAGGAACAGCTATGAC.

## FEATURES

source

1. .475  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone\_lib="NF104H11R"  
 /clone\_lib="Irradiated"  
 /tissue\_type="seedlings"  
 /dev\_stage="seedling"

/note="Vector: Lambda Zap; Seedlings were exposed either to 100 gy gamma or 0.5, 1, 5, or 10 kJ/m2 UV irradiation. Gamma-irradiated samples were harvested at 6, 12, 24 and 48 hours after treatment. UV-irradiated samples were harvested 24 hours post-treatment. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 134 a 71 c 113 g 157 t  
 ORIGIN

Query Match 4.4%; Score 36; DB 14; Length 475;  
 Best Local Similarity 100.0%; Pred. No. 7.7;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 CTTTGTTCCTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
 |||||  
 Db 435 CTTTGTTCCTAAATAAAAAAAAAAAAAAAAAAAAAA 470

## RESULT 5

BH678808/c

LOCUS

DEFINITION BOMB30TF BO\_2\_3\_KB Brassica oleracea genomic clone BOMB30, DNA  
 566 bp DNA linear GSS 19-FEB-2002

ACCESSION

BH678808

VERSION

GSS.

KEYWORDS

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 666)

Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOMB30TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: Tf

Class: sheared ends.

Location/Qualifiers

1. .666

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

## FEATURES

source

/clone="BOMB30"  
 /clone\_lib="BO\_2\_3\_KB"  
 /note="Vector: pHSOI; Site\_1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into pHSOI using BstXI linkers."  
 189 a 137 c 147 g 193 t

## Query Match

Best Local Similarity 100.0%; Pred. No. 5.6;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 GATATCATTCGACGGCAGTTACTAGTCTCTCCCC 124  
 |||||  
 Db 363 GATATCATTCGACGGCAGTTACTAGTCTCTCCCC 328

## RESULT 6

BE209488

LOCUS

DEFINITION

BE209488

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

1 (bases 1 to 240)

Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna  
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
 R., Waterston,R. and Wilson,R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Putative full length read

vector to vector length is This clone is available through: ResGen,  
 Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801  
 For further information call: (800)-533-4363 or contact via email:  
 ccu@resgen.com

Insert length: 647 Std Error: 0.00

High quality sequence stop: 221.

Location/Qualifiers

1. .240

/organism="Glycine max"

/db\_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl037-3361"

/clone\_lib="Gm-cl037"

/tissue\_type="fully expanded leaves of greenhouse grown  
 plants"

/dev\_stage="2 week old"

/lab\_host="DH10B"

/note="vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
 cDNA library was constructed from mRNA isolated from fully  
 expanded leaves of greenhouse grown plants that were 2  
 weeks old. The library was prepared using the Life  
 Technologies pSuperScript cDNA library construction kit.  
 Complementary DNA was synthesized from mRNA using a  
 poly(dT) sequence with a NotI restriction site. SalI  
 linkers adapters were ligated to the blunt-ended cDNA  
 fragments followed by NotI digestion. The cDNA fragments  
 were directionally cloned into the NotI-SalI restriction

## FEATURES

source

site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electro-Max DH10B host cells. This library was constructed in the laboratory of Dr. Lila Vodkin by Anu Khanna at the University of Illinois at Urbana-Champaign. email: l-vodkin@uiuc.edu"

BASE COUNT 137 a 16 c 26 g 47 t 14 others  
ORIGIN  
Query Match 4.3%; Score 35; DB 10; Length 240;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 775 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||

Db 95 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 129  
|||||

RESULT 7  
AI826026/c  
LOCUS  
DEFINITION wk14e08.x1 NCI\_CGAP\_Lym12 Homo sapiens cDNA clone IMAGE:2412326 3',  
mRNA sequence.  
ACCESSION AI826026  
VERSION AI826026.1 GI:5446697  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 284)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Life Technologies catalog #: 11547-015  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.blo.lnlni.gov/dbrrp/image/image.html  
Insert length: 1939 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 271.

FEATURES  
Source  
1..284  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2412326"  
/lab\_host="NCI-CGAP\_Lym12"  
/tissue\_type="lymphoma, follicular mixed small and large  
cell"  
/lab\_host="DH10B"  
/note="Organ: lymph node; Vector: pCMV-SPORT6; Site\_1:  
SalI; Site\_2: NotI; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 1.25 kb. Life Technologies  
catalog #: 11547-015"  
BASE COUNT 102 a 36 c 38 g 108 t  
ORIGIN

Query Match 4.3%; Score 35; DB 9; Length 284;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 775 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||

Db 40 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 6  
|||||

RESULT 8  
AW432208  
LOCUS  
DEFINITION sh70e03.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
292 bp mRNA linear EST 03-DEC-2001

Gm-cl015-5141 5', mRNA sequence.

ACCESSION AW432208  
VERSION AW432208.1 GI:6963515  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max

REFERENCE 1 (bases 1 to 292)  
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna  
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

COMMENT Contact: Shoemaker R/Public Soybean EST Project  
Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
Seq primer: -40RP from Gibco  
High quality sequence stop: 268.

FEATURES  
Source  
1..292  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-5141"  
/tissue\_type="Mature flowers, field grown plants"  
/lab\_host="XL10-Gold"  
/note="Vector: pBluescript II XR; Site\_1: EcoRI; Site\_2:  
XhoI; This cDNA library was constructed from mRNA isolated  
from mature flowers of field grown plants. The cDNA  
library was prepared using the Stratagene pBluescript II  
XR cDNA library construction kit. Complementary DNA was  
synthesized from mRNA using a primer consisting of a poly  
(dT) sequence with a XhoI restriction site. EcoRI adapters  
were ligated to the blunt-ended cDNA fragments followed by  
XhoI digestion. The cDNA fragments were directionally  
cloned into the EcoRI-XhoI restriction site of the  
pBluescript vector. The ligated cDNA fragments were  
transformed into XL10-Gold host cells. This library was  
constructed by Dr. Randy Shoemaker and Dr. John  
Erpelding."

BASE COUNT 162 a 14 c 42 g 72 t 2 others  
ORIGIN  
Query Match 4.3%; Score 35; DB 10; Length 292;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 775 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||

Db 165 TTGTGTTCTAAATAAAAAAAAAAAAAAAAAAAAAA 199  
|||||

RESULT 9  
BI344610  
LOCUS  
DEFINITION BI344610 325 bp mRNA linear EST 30-JUL-2001  
ACCESSION BI344610  
VERSION BI344610.1 GI:15037899  
KEYWORDS EST.  
SOURCE pig.



ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 325)  
AUTHORS Fahrrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
and Keele,J.W.  
TITLE Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACACGCTATGACCAT  
BACKWARD: GTTTCCTCCAGTCACGCG  
Plate: 119 row: L column: 18  
Seq primer: ATTAGCTGACACTATAG.  
Location/Qualifiers  
1..325  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 2pIG"  
/tissue\_type="pooled"  
/lab\_host="DH108"  
/note="Vector: PCMV SPORT6; Site:1: NctI; Site:2: SalI;  
Library made from pooled tissue from testis, ovary,  
endometrium, hypothalamus, pituitary, and placenta."  
BASE COUNT 73 a 100 c 71 g 81 t  
ORIGIN  
Query Match 4.3%; Score 35; DB 13; Length 325;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 775 TTGTGTTCTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 200 TTGTGTTCTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 234  
RESULT 10  
AI268815/c  
LOCUS AI268815  
DEFINITION q04h02.x1 NCI\_CGAP\_Lu5 Homo sapiens cDNA clone IMAGE:1911027 3',  
mRNA sequence.  
ACCESSION AI268815  
VERSION AI268815.1 GI:3887982  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 342)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40Up from Gibco  
High quality sequence stop: 303.  
Location/Qualifiers  
1..342  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:1911027"  
/clone\_lib="NCI\_CGAP\_Lu5"  
/tissue\_type="carcinoid"  
/lab\_host="DH108"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 116 a 50 c 67 g 109 t  
ORIGIN  
Query Match 4.3%; Score 35; DB 9; Length 342;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 775 TTGTGTTCTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 40 TTGTGTTCTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 6  
RESULT 11  
AI957804/c  
LOCUS AI957804  
DEFINITION f06b005.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone  
IMAGE:3730065 3', mRNA sequence.  
ACCESSION AI957804  
VERSION AI957804.1 GI:5750513  
KEYWORDS EST.  
SOURCE zebrafish.  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.  
REFERENCE 1 (bases 1 to 439)  
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy  
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood  
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,  
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,  
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.  
and Willson,R.  
WashU Zebrafish EST Project 1998  
JOURNAL Unpublished (1998)  
COMMENT Contact: Stephen L. Johnson  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: zbrfish@watson.wustl.edu  
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
Matthew Clark. DNA Sequencing by: Washington University Genome  
Sequencing Center Clone distribution: Genome Systems, St. Louis,  
Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
[Resourcen2@centrumprimaridatenbank.de](mailto:Resourcen2@centrumprimaridatenbank.de)  
Seq primer: T7 ET from Amersham  
High quality sequence stop: 356.  
Location/Qualifiers  
1..439  
/organism="Danio rerio"  
/db\_xref="taxon:7955"  
/clone="IMAGE:3730065"

/clone\_lib="zebrafish WashU MPMG EST"  
 /sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield stage embryos"  
 /lab\_host="XLI-blue MRF"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5'pGACTAGTCTTAGATCGGAGCGGCCCTTTTCTTTTCTTTT3']; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."

BASE COUNT 163 a 57 c 59 g 160 t

## ORIGIN

Query Match 4.3%; Score 35; DB 9; Length 439;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
 ||||||||||||||||||||||||||||||||||||  
 Db 36 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 2

## RESULT 12

AJ500337/c

## LOCUS

DEFINITION AJ500337 MTGIM Medicago truncatula cDNA clone mtgmacc120015c09,  
 mRNA sequence.

## ACCESSION

AJ500337

## VERSION

AJ500337.1 GI:22081270

## KEYWORDS

EST.

## SOURCE

barrel medic.

## ORGANISM

Medicago truncatula

## REFERENCE

1 (bases 1 to 542)

## AUTHORS

Krajinski, F., Manthey, K., Bartelsmeier, V., Meyer, F., Bartels, D.,  
 Bekel, T., Linke, B., Franken, P., Kuester, H., Perlick, A. M. and  
 Puehler, A.

## TITLE

Detection of transcript sequences from mycorrhizal roots of the  
 model mycorrhiza Medicago truncatula genotype A17 - Glomus  
 intraradices using the approach of an EST genome project based on  
 an SSH library

## JOURNAL

Unpublished (2002)

## COMMENT

Contact: Krajinski F

LG Molekulargenetik

Herrenhaeuser Str. 2 D-30419 Hannover, Germany.

Location/Qualifiers

1..542

/organism="Medicago truncatula"

/db\_xref="taxon:3880"

/clone="mtgmacc120015c09"

/clone\_lib="MTGIM"

/tissue\_type="mycorrhizal roots"

/dev\_stage="3 weeks after  
 inoculation"

/note="Vector: pGEM-Teasy; genotype A17; cDNA was prepared  
 from total RNA using the SMART PCR cDNA system (Clontech)  
 from roots harvested three weeks after inoculation with  
 Glomus intraradices. This cDNA was used as tester in a

Suppression Subtractive Hybridization (SSH). The SSH-cDNA  
 fragments were generated using the SSH-adaptor sequences  
 ctaatcagctactataggctcagcgccgcgcgcgcaggt and  
 ctaatcagctactataggctcagcgccgcgcgcgcaggt (Clontech) and  
 ligated after Suppression Subtractive Hybridization in to  
 the pGEM-Teasy vector from Promega. Plasmids containing  
 cDNA inserts were propagated in E. coli TOP 10F' cells  
 (Invitrogen)"

BASE COUNT 173 a 98 c 91 g 180 t

## ORIGIN

Query Match 4.3%; Score 35; DB 9; Length 542;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
 ||||||||||||||||||||||||||||||||||||  
 Db 71 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 37

## RESULT 13

AL629881

## LOCUS

DEFINITION AL629881 XGC-gastrula Silurana tropicalis cDNA clone TGas024n16 5',  
 mRNA sequence.

## ACCESSION

AL629881

## VERSION

AL629881.1 GI:16599364

## KEYWORDS

EST.

## SOURCE

western clawed frog.

## ORGANISM

Silurana tropicalis

## REFERENCE

1 (bases 1 to 575)

## AUTHORS

Huckie, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.

## TITLE

Sanger Xenopus tropicalis EST project 2001 (10\_2001)

## JOURNAL

Unpublished (2001)

## COMMENT

Contact: Huckie E

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas024n16.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1..575

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="TGas024n16"

/clone\_lib="XGC-gastrula"

/dev\_stage="gastrula (stages 10.5-13 mixed)"

/lab\_host="Escherichia coli XLI-blue"

/note="Vector: pCSI07; Site\_1: EcoRI; Site\_2: NotI; cDNA  
 was oligo dt primed from 5ug of poly A+ RNA from stages  
 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
 into pCSI07 with EcoRI at the 5' end and NotI at the 3'  
 end."

BASE COUNT 163 a 123 c 144 g 145 t

## ORIGIN

Query Match 4.3%; Score 35; DB 9; Length 575;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
 ||||||||||||||||||||||||||||||||||||  
 Db 535 TTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 569

## RESULT 14

AA584642/c

```

LOCUS      AA584642      135 bp      mRNA      linear      EST 26-SEP-1997
DEFINITION no12f10.s1 NCI_CGAP_Phel Homo sapiens cDNA clone IMAGE:1100491 3',
similar to gb:M13452 LAMIN A (HUMAN)), mRNA sequence.
ACCESSION  AA584642
VERSION    AA584642.1 GI:2369251
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 135)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: W. Marston Linehan, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: Stratagene, Inc.
            DNA Sequencing by: Greg Lennon, Ph.D.
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/dbbrp/image/image.html
            Insert Length: 188 Std Error: 0.00
            Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES   Location/Qualifiers
            source          1..135
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:1100491"
                        /clone_lib="NCI_CGAP_Phel"
                        /tissue_type="pheochromocytoma"
                        /lab_host="SOLR (kanamycin resistant)"
                        /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                        : Cloned unidirectionally. Primer: Oligo dT.
                        Pheochromocytoma, 5' adaptor sequence: 5' GAATTCGGGACGAG
                        3' 3' adaptor sequence: 5' CTCGAGTTTTCCTTTTTCCTTTT 3'."
                        Average insert size: 1.3 kb."
BASE COUNT      30 a      19 c      34 g      52 t
ORIGIN
Query Match      4.2%; Score 34; DB 9; Length 135;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
      |||||||||||||||||||||||||||||||||||
Db 40 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 7

RESULT 15
AW101025
LOCUS      AW101025      202 bp      mRNA      linear      EST 30-NOV-2001
DEFINITION sd64f12.v1 Gm-cl008 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl008-1128 5', mRNA sequence.
ACCESSION  AW101025
VERSION    AW101025.1 GI:6071638
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
            1 (bases 1 to 202)
REFERENCE  Shoemaker, R., Keim, P., Vodkin, L., Erpellding, J., Coryell, V., Khanna
            , A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
            Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
            , Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
            , R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
            , R., Waterston, R. and Wilson, R.

```

```

TITLE      Public Soybean EST Project
JOURNAL    Unpublished (1999)
COMMENT    Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Putative full length read
            vector to vector length is 706 This clone is available through:
            Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
            35801 For further information call: (800)-533-4363 or contact via
            email: ccu@resgen.com
            Seq primer: -40Rp from Gibco
            High quality sequence stop: 139.
FEATURES   Location/Qualifiers
            source          1..202
                        /organism="Glycine max"
                        /db_xref="taxon:3847"
                        /clone="GENOME SYSTEMS CLONE ID: Gm-cl008-1128"
                        /clone_lib="Gm-cl008"
                        /lab_host="DHI0B"
                        /note="vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                        cDNA library was constructed from mRNA isolated from whole
                        young pods, approximately 2cm long, of 12-week-old
                        greenhouse grown plants. The library was prepared using
                        the Life Technologies pSuperScript cDNA library
                        construction kit. Complementary DNA was synthesized from
                        mRNA using a poly (dT) sequence with a NotI restriction
                        site. SalI linkers adapters were ligated to the
                        blunt-ended cDNA fragments followed by NotI digestion. The
                        cDNA fragments were directionally cloned into the
                        NotI-SalI restriction site of the pSPORT1 vector. The
                        ligated cDNA fragments were transformed into E. coli
                        Electromax DH10B host cells. This library was constructed
                        by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT      132 a      15 c      12 g      43 t
ORIGIN
Query Match      4.2%; Score 34; DB 10; Length 202;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
      |||||||||||||||||||||||||||||||||||
Db 100 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 133

Search completed: January 25, 2003, 18:17:40
Job time : 2200 secs

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 01:26:02 ; Search time 265 Seconds

(without alignments)

6874.968 Million cell updates/sec

Title: US-09-822-080B-1

Perfect score: 809

Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	240.6	29.7	792	24	ABN99075
C 2	50	6.2	8876	24	Arabidopsis thalia
C 3	49.2	6.1	385	22	Human immune syste
C 4	48.4	6.0	1255	19	Human immune/haema
C 5	48	5.9	6699	24	Human secreted pro
C 6	46.8	5.8	13377	22	Human chemically m
C 7	46.8	5.8	13377	24	Tumour suppressor
C 8	46	5.7	6255	24	Human immune syste
C 9	45.8	5.7	2206	23	Human immune syste
					DNA encoding novel

C 10	45.6	5.6	281	22	AA08167	Human breast cance
C 11	45.6	5.6	855	22	AAH34676	Human colon cancer
C 12	45.6	5.6	1012	23	ABV25209	Human prostate exp
C 13	45.6	5.6	1012	23	ABV25860	Human prostate exp
C 14	45.6	5.6	5908	22	AA545386	Chemically pretrea
C 15	45.6	5.6	5908	24	AA561216	Human gene regulat
C 16	45.6	5.6	5908	24	ABK28231	DNA transcription
C 17	45.6	5.6	11745	24	ABK28331	DNA transcription
C 18	45.6	5.6	5891	24	AA561088	Human gene regulat
C 19	45.4	5.6	5942	24	ABN80031	Human chemically m
C 20	45.4	5.6	7810	22	AA545435	Chemically pretrea
C 21	45.4	5.6	7810	24	ABK28282	DNA transcription
C 22	45.2	5.6	6409	22	AA545496	Tumour suppressor
C 23	45	5.6	1817	24	AA033245	Human secreted pro
C 24	44.6	5.5	635	23	AA565555	DNA encoding novel
C 25	44.6	5.5	916	13	AAQ20375	Sequence encoding
C 26	44.6	5.5	1715	24	ABQ54175	Human ovarian anti
C 27	44.6	5.5	6216	24	ABL70139	Human chemically p
C 28	44.6	5.5	6216	24	ABK39932	Human chemically p
C 29	44.4	5.5	426	22	AA188844	Human polynucleoti
C 30	44.4	5.5	761	24	ABN99190	Arabidopsis thalia
C 31	44.4	5.5	1052	10	AA950224	Malaria-specific D
C 32	44.4	5.5	1450	21	AA558816	DNA encoding a hum
C 33	44.4	5.5	2498	21	AA27988	Wheat CCR4 transcr
C 34	44.4	5.5	6665	22	AA545298	Chemically pretrea
C 35	44.4	5.5	6665	24	ABL32082	Human immune syste
C 36	44.4	5.5	6665	24	ABK28129	DNA transcription
C 37	44.4	5.5	8666	22	AA546306	Tumour suppressor
C 38	44.4	5.5	8666	24	ABQ67178	Human angiogenesis
C 39	44.4	5.5	8666	24	ABK34009	Human DNA for stag
C 40	44.4	5.5	8666	24	ABL32397	Human immune syste
C 41	44.2	5.5	1673	20	AA777320	Human secreted pro
C 42	44	5.4	309	22	AA08549	Human breast cance
C 43	44	5.4	5750	22	AA546708	Tumour suppressor
C 44	44	5.4	5750	24	ABL34008	Human immune syste
C 45	44	5.4	6041	24	ABL32045	Human immune syste

#### ALIGNMENTS

#### RESULT 1

ABN99075

ID ABN99075 standard; DNA; 792 BP.

XX AC ABN99075;

XX DT 01-AUG-2002 (first entry)

XX DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 843.

XX KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress; disease; crop; thale cress; tolerance factor; insect; pathogen; nutrition; ds.

XX OS Arabidopsis thaliana.

XX PN US2002023281-A1.

XX PD 21-FEB-2002.

XX PF 26-JAN-2001; 2001US-0770445.

XX PR 27-JAN-2000; 2000US-178472P.

XX PA (GORL/) GORLACH J.

XX PA (ANY/) AN Y.

XX PA (HAMI/) HAMILTON C M.

XX PA (PRIC/) PRICE J L.

XX PA (RAIN/) RAINES T M.

XX PA (YUY/) YU Y.

XX PA (RAME/) RAMEAKA J G.

XX PA (PAGE/) PAGE A.









```

PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.

PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057659.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
XX Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
XX Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
XX Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
XX WPI; 1998-506364/43.
XX P-PSDB; AAW74795.
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
XX - useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX
XX Claim 1; Page 297-298; 721pp; English.
XX
XX This sequence represents a nucleic acid molecule designated Gene 65 from
XX the human cDNA clone HSGBA84 (deposited as clone ATCC 97900 and ATCC
XX 209046) which encodes a secreted human protein. The gene can be used to
XX generate fusion proteins by linking to the gene to a human
XX immunoglobulin FC portion (e.g. AAV59502) for increasing the stability of
XX the fused protein as compared to the human protein only.
XX The invention relates to 186 novel genes and their fragments (nucleic
XX acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
XX which are useful for preventing, treating or ameliorating medical
XX conditions e.g. by protein or gene therapy. Also, pathological conditions
XX can be diagnosed by determining the amount of the new polypeptides in a
XX sample or by determining the presence of mutations in the new
XX polynucleotides. Specific uses are described for each of the 186
XX polynucleotides, based on which tissues they are most highly expressed in
XX (see AAV59511 for described uses).
XX
XX Sequence 1255 BP; 324 A; 297 C; 240 G; 391 T; 3 other;
XX
Query Match 6.0%; Score 48.4; DB 19; Length 1255;
Best Local Similarity 53.1%; Pred. No. 0.039;
Matches 103; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 616 CACTTCATCCAAGACTATGCTCTATTATCTGAGAGAAATTAAGACACCTTAATAAGAGG 675
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1030 CACATTATAATCCCACTCGCTCTTTTTCATGCCACATAATAACVACGCAATATGG 1089
QY 676 ATAGTGTTTATACTTACCTCTTAATAATAAACTCTATCTATGATGTTTCTTTGT 735
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1090 ATACGCCTTATTGATTAACTAGTTCCTAATGATGGACTTTTAAGTTGTTTCTTTT 1149
QY 736 TCATCGATCATCATGATGATGAATAAACAATCTTCCTTTGTTTCTAAAAA 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1150 TTTTCTTTTCTGCTACTGCAACGATGTATAATAATGTCCTTATCAAAAAA 1209
QY 796 AAAAAA 809
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1210 AAAAAA 1223
XX
XX RESULT 5
XX ABN80183/c
XX ID ABN80183 standard; DNA; 6699 BP.
XX
XX AC ABN80183;
XX
XX DT 15-JUL-2002 (first entry)
XX

```

```
DE Human chemically modified disease associated gene SEQ ID NO 200.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07536.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130908/17.
XX
XX Novel nucleic acid useful for diagnosis and therapy of diseases
XX associated with development genes such as diabetes, comprises a
XX sequence of a segment of chemically pretreated DNA of genes associated
XX with development.
XX
XX Claim 1; SEQ ID NO 200; 27pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
XX least 18 bases in length of a segment of chemically pretreated DNA (II)
XX of genes associated with development selected from 87 genes listed in
XX the specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
XX sequences (ABN79984-ABN80333) or their complements. The invention is
XX useful for the diagnosis or therapy of diseases associated with
XX development genes in particular disease related to homeobox containing
XX genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
XX associated with congenital heart disease, epilepsy, diseases related to
XX histone deacetylation, Curarino syndrome, diseases related with the
XX development of the brain and limb girdle muscular dystrophy and dwarfism.
XX Oligomers specific to each of the genes are useful for detecting the
XX methylation state of all CpG dinucleotides within the 350 sequences or
XX (II) and their complementary sequences, as primer oligonucleotides for
XX the amplification of the 350 sequences, (II) and/or their complements and
XX as oligomer probes for detecting the cytosine methylation state and/or
XX single nucleotide polymorphisms (SNPs).
XX Note: The sequence data for this patent did not form part of the printed
XX specification but is based on sequence information supplied to Derwent by
XX the European Patent Office.
XX
XX Sequence 6699 BP; 1535 A; 294 C; 1763 G; 3107 T; 0 other;
XX
XX Query Match 5.9%; Score 48; DB 24; Length 6699;
XX Best Local Similarity 56.2%; Pred. No. 0.081;
XX Matches 90; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
XX
QY 650 AATAAAGACCCTTAATAAGAGGTAAGTGTATAACTTACCTCTAATAATAAACT 709
DB 5372 ACATTAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 5313
QY 710 CTATCTATGATGATGTTTCTTGTTCATCATCATCATCATCATCATCATCATCAT 769
DB 5312 CTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 5253
QY 770 TTTCTTGTCTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 809
DB 5252 ATTATATATCTATCAATAATAATAATAATAATAATAATAATAATAATAATA 5213
XX
XX RESULT 6
```

```
AAS46475/C
ID AAS46475 standard; DNA: 13377 BP.
XX
AC AAS46475;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #197.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX analysing diseases associated with cytosine methylation state e.g.
XX cancer.
XX
XX Claim 1; SEQ ID No 197; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisphosphate, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (SS) and sequences complementary to (SS). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 13377 BP; 3915 A; 146 C; 3046 G; 6270 T; 0 other;
XX
XX Query Match 5.8%; Score 46.8; DB 22; Length 13377;
XX Best Local Similarity 55.6%; Pred. No. 0.2;
XX Matches 90; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
XX
QY 646 AGAGAAATTAAGACCACTTAATAAGAGGATAAGTGTATTAACCTTACCTTAATAATA 705
DB 12200 ATACTAACGATAATACACTTAATATTCATAAACCCTTTTCGCTTATATATATATA 12141
```



Db 5972 ACTAAAAATAAAA 5959  
RESULT 9  
ID ABK43595 standard; cDNA; 2206 BP.  
XX AC ABK43595;  
XX DT 05-JUN-2002 (first entry)  
XX DE DNA encoding novel central nervous system protein #175.  
XX KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiodysplasia;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy; gene; ss.  
XX KW Homo sapiens.  
XX OS  
XX PN WO20015318-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01332.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234597.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239835.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241321.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.





ABV25860  
ID ABV25860 standard; cDNA; 1012 BP.  
XX AC  
XX ABV25860;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 25851.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX Homo sapiens.  
XX  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 5217; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 1012 BP; 274 A; 228 C; 219 G; 286 T; 5 other;  
XX  
Query Match 5.6%; Score 45.6; DB 23; Length 1012;  
Best Local Similarity 54.1%; Pred. No. 0.18; Mismatches 79; Indels 0; Gaps 0;  
Matches 93; Conservative 0;  
QY 638 ATTATCTGAGAGAAATTAAGACCACTTAAATAAGAGGATAAGTGTATTAACCTTACCTCT 697  
DB 771 ATAAATTTGAACAAATTTAATACACTTCCCTCATGCCCTTCTATATATAAACAATTAATACC 830  
QY 698 AATAAATAACCTATCTATGATGATGTTTCTTTGTTTCATCGATCAATCATGATG 757  
DB 831 ATTAGTTCCTCCATCTTGACATTTTATTTTCAGTTTTTATATATATATTTTGAATATT 890  
QY 758 GAATAAACAATCTTCTCTTTGTTTCTTAAATAAAAAAAAAAAAAAAAAAAAAA 809  
DB 891 TATTAATTTATCTGACCTACAGAACTTAAATAAAAAAAAAAAAAAAAAAAAAA 942

RESULT 14  
AAS45386/C  
ID AAS45386 standard; DNA; 5908 BP.  
XX  
XX AAS45386;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
XX Chemically pretreated genomic DNA associated with cell cycle #46.  
XX  
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;  
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
KW PCR primer.  
XX  
XX Homo sapiens.  
XX  
XX WO200168911-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 15-MAR-2001; 2001WO-EP02945.  
XX  
XX 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-602751/68.  
XX  
XX Designing primers and probes for analysing diseases associated with  
PT cytosine methylation state e.g. arthritis, cancer, aging,  
PT arteriosclerosis comprising fragments of chemically modified genes  
PT associated with cell cycle -  
XX  
XX Claim 1; SEQ ID No 91; 28pp; English.  
XX  
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
CC molecules associated with the cell cycle and specific PCR primers of the  
CC invention. The sequences are useful for detecting the methylation state  
CC of all CpG dinucleotides in a sequence and therefore for analysing  
CC associated diseases. By analysing cytosine methylations in the pretreated  
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
CC of existing diseases or the predisposition to specific diseases can be  
CC ascertained. The parameters may be compared to another set of genetic  
CC and/or epigenetic parameters, the differences serving as basis for  
CC diagnosis and/or prognosis events which are disadvantageous to patients.  
CC The sequences of the invention are useful for the diagnosis and therapy  
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
CC aging, glomerular disease, Lewy body disease, arthritis,  
CC arteriosclerosis, solid tumours and cancers.  
XX  
XX Sequence 5908 BP; 1454 A; 149 C; 1327 G; 2978 T; 0 other;  
XX  
Query Match 5.6%; Score 45.6; DB 22; Length 5908;  
Best Local Similarity 54.1%; Pred. No. 0.3;  
Matches 93; Conservative 0; Mismatches 79; Indels 0; Gaps 0;  
QY 633 TGCTATTATCTGAGAAATTAAGACCACTTAAATAAGAGGATAAGTGTATTAACCTTA 692  
DB 1658 TATATAAATA 1599  
QY 693 CCTCTAATAAACAATCTATCTATGATGATGTTTCTTTGTTTCATCGATCATCAT 752  
DB 1598 TTAAT 1539







EARLIER APPLICATION NUMBER:	60/047,500
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,587
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,492
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,598
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,613
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,582
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,596
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,612
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,632
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,601
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/043,580
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,568
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,314
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,569
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,311
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,671
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,674
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,669
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,312
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,313
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,672
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043,315
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/048,974
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/056,886
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,877
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,889
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,662
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,893
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,630
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,878
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,637
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,903
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,888
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,879
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056,880

1	EARLIER FILING DATE:	1997-08-22
2	EARLIER APPLICATION NUMBER:	60/056,894
3	EARLIER FILING DATE:	1997-08-22
4	EARLIER APPLICATION NUMBER:	60/056,911
5	EARLIER FILING DATE:	1997-08-22
6	EARLIER APPLICATION NUMBER:	60/056,636
7	EARLIER FILING DATE:	1997-08-22
8	EARLIER APPLICATION NUMBER:	60/056,874
9	EARLIER FILING DATE:	1997-08-22
10	EARLIER APPLICATION NUMBER:	60/056,910
11	EARLIER FILING DATE:	1997-08-22
12	EARLIER APPLICATION NUMBER:	60/056,864
13	EARLIER FILING DATE:	1997-08-22
14	EARLIER APPLICATION NUMBER:	60/056,631
15	EARLIER FILING DATE:	1997-08-22
16	EARLIER APPLICATION NUMBER:	60/056,845
17	EARLIER FILING DATE:	1997-08-22
18	EARLIER APPLICATION NUMBER:	60/056,892
19	EARLIER FILING DATE:	1997-08-22
20	EARLIER APPLICATION NUMBER:	60/057,761
21	EARLIER FILING DATE:	1997-08-22
22	EARLIER APPLICATION NUMBER:	60/047,595
23	EARLIER FILING DATE:	1997-05-23
24	EARLIER APPLICATION NUMBER:	60/047,599
25	EARLIER FILING DATE:	1997-05-23
26	EARLIER APPLICATION NUMBER:	60/047,588
27	EARLIER FILING DATE:	1997-05-23
28	EARLIER APPLICATION NUMBER:	60/047,585
29	EARLIER FILING DATE:	1997-05-23
30	EARLIER APPLICATION NUMBER:	60/047,586
31	EARLIER FILING DATE:	1997-05-23
32	EARLIER APPLICATION NUMBER:	60/047,590
33	EARLIER FILING DATE:	1997-05-23
34	EARLIER APPLICATION NUMBER:	60/047,594
35	EARLIER FILING DATE:	1997-05-23
36	EARLIER APPLICATION NUMBER:	60/047,589
37	EARLIER FILING DATE:	1997-05-23
38	EARLIER APPLICATION NUMBER:	60/047,593
39	EARLIER FILING DATE:	1997-05-23
40	EARLIER APPLICATION NUMBER:	60/047,614
41	EARLIER FILING DATE:	1997-05-23
42	EARLIER APPLICATION NUMBER:	60/043,578
43	EARLIER FILING DATE:	1997-04-11
44	EARLIER APPLICATION NUMBER:	60/043,576
45	EARLIER FILING DATE:	1997-04-11
46	EARLIER APPLICATION NUMBER:	60/047,501
47	EARLIER FILING DATE:	1997-05-23
48	EARLIER APPLICATION NUMBER:	60/043,670
49	EARLIER FILING DATE:	1997-04-11
50	EARLIER APPLICATION NUMBER:	60/056,632
51	EARLIER FILING DATE:	1997-08-22
52	EARLIER APPLICATION NUMBER:	60/056,664
53	EARLIER FILING DATE:	1997-08-22
54	EARLIER APPLICATION NUMBER:	60/056,876
55	EARLIER FILING DATE:	1997-08-22
56	EARLIER APPLICATION NUMBER:	60/056,881
57	EARLIER FILING DATE:	1997-08-22
58	EARLIER APPLICATION NUMBER:	60/056,909
59	EARLIER FILING DATE:	1997-08-22
60	EARLIER APPLICATION NUMBER:	60/056,908
61	EARLIER FILING DATE:	1997-08-22
62	EARLIER APPLICATION NUMBER:	60/048,964
63	EARLIER FILING DATE:	1997-06-06
64	EARLIER APPLICATION NUMBER:	60/057,650
65	EARLIER FILING DATE:	1997-09-05
66	EARLIER APPLICATION NUMBER:	60/056,884
67	EARLIER FILING DATE:	1997-08-22



```

724   TGTTTTCTTTTGTTTCATCATCATCATCATGTAATAAAACATCTTCTCCTTGTCTTCT 783
Db    1163 TGTCCTTTTAGAATGTAATGATGATAAGTTGTGAATAAAGGTTTCTATCTAGTTGT 1222

784   AAAAAAAAAAAAAAAAAAAAAAAA 808
Db    1223 AAAAAAAAAAAAAAAAAAAAAAAA 1247

RESULT 5
; JS-08-993-228-20
; Sequence 20, Application US/08993228
; Patent No. 5976838
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,228
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-993-228-20

Query Match          5.2%; Score 41.8; DB 2; Length 3113;
Best Local Similarity 63.4%; Pred No. 0.087;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps

QY      706   AACCTATCTATGATGATGTTTTCTTTCTTCATCGATCATCATGTAAGAATAAAA 765
Db       3013 AAGCCTTTCCATCATCTCTTCTATTAAAAAGAATGTGATTTATTGTTTTTATGAATAAAA 3077

QY      766   CATCTTTCTTGGTTCTAAAAAATAAAAAAAAAAAAAAAAAAAAA 806
Db       3073 TATACCTCGTGTGTTCTCAAAAAATAAAAAAAAAAAAAAA 3113

RESULT 6
US-08-545-196B-10
Application US/08545196B

```



APPLICANT: Giri, Judith G  
TITLE OF INVENTION: Interleukin-15 Receptors  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple Operating System 7.1  
SOFTWARE: Microsoft Word for Apple, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/300,903A  
FILING DATE: 06-SEPTEMBER-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN 08/236,919  
FILING DATE: 06-MAY-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Perkins, Patricia Anne  
REGISTRATION NUMBER: 34,695  
REFERENCE/DOCKET NUMBER: 2822-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-587-0430  
TELEFAX: 206-233-0644  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1641 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..839  
US-08-300-903A-8  
Query Match 5.0%; Score 40.6; DB 1; Length 1641;  
Best Local Similarity 62.1%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;  
Matches 64; Conservative 0  
Qy 707 ACTCTATCTATGATGATGTTTCTTTGTTTCATCGATCATCATCATGATGATGGAATAAAC 766  
Db 1475 ATTATATCTATACATTATATATTTGTTGACATTTGACCTTGTGATAAACAAAAAT 1534  
Qy 767 ATCTTCTCTTTGTTCTTCAAAAAAATAAAAAAATAAAAAA 809  
Db 1535 AAAACATCTATTTTCAATAAAAAAATAAAAAAATAAAAAA 1577  
RESULT 11  
US-08-749-522-4  
Sequence 4, Application US/08749522  
Patent No. 6096950  
GENERAL INFORMATION:  
APPLICANT: John, Maliyakal  
TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: WI  
COUNTRY: U.S.A.  
ZIP: 53202-4497  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/749,522  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Baker, Jean C. 35,433  
REGISTRATION NUMBER: 670513.90244  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (414) 277-5709  
TELEFAX: (414) 271-3552  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2781 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-749-522-4  
Query Match 5.0%; Score 40.6; DB 3; Length 2781;  
Best Local Similarity 51.4%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;  
Matches 94; Conservative 0  
Qy 626 AAGACTATCTCTATTTATCTGAGAGAAATTAAGACCACTTAATAAGAGGATAAGTGTTA 685  
Db 1118 AAAATTTTATGCAAGATGATATATATAAATAAATTAATAATATATATATTAATA 1177  
Qy 686 TAACTTACCTCTAATAATAAACTCTATATATGATGATGTTTCTTTGTTTCATCGATCA 745  
Db 1178 TAATATAAAATAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1237  
Qy 746 TCATCATGATGATGAATAAACATCTTCTCTTTGTTTCTTAAATAAAAAAATAAAAAA 805  
Db 1238 TTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1297  
Qy 806 AAA 808  
Db 1298 TAA 1300  
RESULT 12  
US-08-123-934A-5  
Sequence 5, Application US/08123934A  
Patent No. 6291206  
GENERAL INFORMATION:  
APPLICANT: WOZNEY, John  
APPLICANT: CELESTE, Anthony J.  
APPLICANT: THIES, R. Scott  
APPLICANT: YAMAJI, No. 6291206oru  
TITLE OF INVENTION: RECEPTOR PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute Inc. - Legal Affairs  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: USA  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/123,934A  
FILING DATE: 17-SEP-1993  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:

```
; NAME: LAZAR, Steven R
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 876 1170
; TELEFAX: 617 876 5851
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: CFK1-10a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 474..2000
; PCT-US94-10080-5
; US-08-123-934A-5

Query Match          4.9%; Score 40; DB 4; Length 3238;
Best Local Similarity 54.9%; Pred. No. 0.26;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 666 AATAAGAGGATAAGTGTATTAACCTCTCTAATAATAAAACCTCTATCTATGTATGATG 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3046 ACTATCCATTACATGCATTAAACTCTGCCAGAAAAAATAAATACTATTTGTTTAAATC 3105

QY 726 TTTTCTTTGTCATCGATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3106 TACTTTTGTATTAGTAGTATTATTGTATATAAATAAATAAATAAATAAATAAATAA 3165

QY 786 AAAAAAAAAAAAAAAAAAAAAA 809
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3166 AAAAAAAAAAAAAAAAAAAAAA 3189

RESULT 13
PCT-US94-10080-5
; Sequence 5, Application PC/TUS9410080
; GENERAL INFORMATION:
; APPLICANT: GENETICS INSTITUTE, INC.
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute Inc.- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10080
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,934
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8260
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 3238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: CFK1-10a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 474..2000
; PCT-US94-10080-5

Query Match          4.9%; Score 40; DB 5; Length 3238;
Best Local Similarity 54.9%; Pred. No. 0.26;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 666 AATAAGAGGATAAGTGTATTAACCTCTCTAATAATAAAACCTCTATCTATGTATGATG 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3046 ACTATCCATTACATGCATTAAACTCTGCCAGAAAAAATAAATACTATTTGTTTAAATC 3105

QY 726 TTTTCTTTGTCATCGATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3106 TACTTTTGTATTAGTAGTATTATTGTATATAAATAAATAAATAAATAAATAAATAA 3165

QY 786 AAAAAAAAAAAAAAAAAAAAAA 809
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3166 AAAAAAAAAAAAAAAAAAAAAA 3189

RESULT 14
US-08-938-675A-1
; Sequence 1, Application US/08938675A
; Patent No. 6107287
; GENERAL INFORMATION:
; APPLICANT: de Lumen, Benito O.
; APPLICANT: Galvez, Alfredo F.
; TITLE OF INVENTION: Lunasin Peptides
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,675A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B98-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 770 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-938-675A-1

Query Match          4.9%; Score 39.4; DB 3; Length 770;
Best Local Similarity 62.9%; Pred. No. 0.23;
Matches 61; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 08:52:08 ; Search time 50 Seconds  
(without alignments)  
7269.105 Million cell updates/sec

Title: US-09-822-080B-1

Perfect score: 809

Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA.:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PTCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	240.6	29.7	792	10	US-09-770-445-843
2	44.4	5.5	157	10	US-09-924-035A-577
3	44.4	5.5	761	10	US-09-770-445-958
4	44.2	5.5	1673	9	US-10-114-893-191
5	44	5.4	477	9	US-09-938-842A-3890
6	43.8	5.4	376	10	US-09-880-107-549
7	43.8	5.4	3725	10	US-09-886-562-6
8	43.6	5.4	759	10	US-09-770-445-972
9	43.2	5.3	295	10	US-09-764-846-33
10	43.2	5.3	304	10	US-09-764-846-109
11	43	5.3	275	10	US-09-878-574-15072
12	43	5.3	298	10	US-09-960-352-1004
13	42.8	5.3	1891	9	US-09-938-842A-5251
14	42.8	5.3	6381	10	US-09-969-347-216
15	42.6	5.3	2500	10	US-09-954-456-724
16	42.6	5.3	2500	10	US-09-954-456-1169
17	42.6	5.3	2500	10	US-09-954-456-1827
18	42.2	5.2	417	10	US-09-924-035A-250
19	42.2	5.2	419	10	US-09-960-352-11234

Sequence 46, Appl  
Sequence 307, Appl  
Sequence 13, Appl  
Sequence 1, Appl  
Sequence 9974, Ap  
Sequence 14101, A  
Sequence 494, App  
Sequence 1215, Ap  
Sequence 15014, A  
Sequence 66, Appl  
Sequence 38, Appl  
Sequence 112, App  
Sequence 153, App  
Sequence 153, App  
Sequence 177, App  
Sequence 23, Appl  
Sequence 224, App  
Sequence 430, App  
Sequence 13, Appl  
Sequence 403, App  
Sequence 130, App  
Sequence 558, App  
Sequence 3061, Ap  
Sequence 351, App

#### ALIGNMENTS

RESULT 1

US-09-770-445-843  
; Sequence 843, Application US/09770445  
; Patent No. US20020023281A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jörn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Krickler, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurbán, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; TITLE OF INVENTION: thaliana  
; FILE REFERENCE: 2023US (PARA-012PRV)  
; CURRENT APPLICATION NUMBER: US/09770.445  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 60/178,472  
; NUMBER OF SEQ ID NOS: 999  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 843  
; LENGTH: 792  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-770-445-843

Query Match 29.7%; Score 240.6; DB 10; Length 792;  
Best Local Similarity 65.6%; Pred. No. 2.3e+46;  
Matches 418; Conservative 0; Mismatches 204; Indels 15; Gaps 4;







Db 161 TTGGAATGATGCATATAATAAATTTTACTGTTTTTAAAAA 220  
QY 802 AAAAAAA 809  
Db 221 AAAAAAA 228

RESULT 11  
US-09-878-574-15072  
; Sequence 15072, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 15072  
; LENGTH: 275  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701069617H1  
US-09-878-574-15072

Query Match 5.3%; Score 43; DB 10; Length 275;  
Best Local Similarity 54.4%; Pred. No. 0.38;  
Matches 130; Conservative 0; Mismatches 105; Indels 4; Gaps 2;

QY 7 TCTATGTTTACTCCTCTTGCT-TTACCACGTGTTTGGCCGCGACCGAAGCGTG 65  
Db 25 TACTATATCCCTTCCCTTCTCTCTCTTTGCTTCTTCTGCTCTTTTCAAGAGATGTTG 84

QY 66 GACCAAGTTCTGCACACTGATGCTGATATCATATTC---GACGGCAGTTACTACGTTCTCC 122  
Db 85 AACAAAGTTCTGCACATAAGTGCGACACCCCATTTTCCAGGTGGCACATATACATATGC 144

QY 123 CCTCATCTGGGCGCTACAGTGGCGGCTAACTCTCGTCTCCGTCGTGGCAACCAAGT 182  
Db 145 CATCAACTTGGGGCGCTGCGGTGGTGGATGACACTAGGCGGACGAACTCAAACT 204

QY 183 GTCCCTCTTTATCGACAGAGCGCTTCAGAGTCAACAGGGGCAATCCCGTGAATTC 241  
Db 205 GCCCAGTTACTGTTTGAAGATTACTCAGAAATCTTCCGTGGCACACAGTCAAAATTC 263

RESULT 12  
US-09-960-352-1004/c  
; Sequence 1004, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 1004  
; LENGTH: 298  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 05-LTB3057-009-Q1-K1-B1  
US-09-960-352-1004

Query Match 5.3%; Score 43; DB 10; Length 298;  
Best Local Similarity 62.6%; Pred. No. 0.39;  
Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 703 TAAACTCTATCTATGATGATGTTTCTTGTTCATCGCATCATCATCATGATGGAATA 762  
Db 186 TAAAGTTTTTTTTTCCCTTTTTTTTTTTTAAATAATTTTTTATTTTTTAAAA 127

QY 763 AACATCTTCTCTGTTTCTAAAAA 809  
Db 126 ATTTAAATTTTAAATTTAAAAA 80

RESULT 13  
US-09-938-842A-5251/c  
; Sequence 5251, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 5251  
; LENGTH: 1891  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-5251

Query Match 5.3%; Score 42.8; DB 9; Length 1891;  
Best Local Similarity 51.0%; Pred. No. 0.81;  
Matches 101; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 611 ACAGACACTTCATCCAGACATGCTCTATTATCTGAGAGAAATTAAGACCCTTAATA 670  
Db 928 ATAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 869

QY 671 AGAGGATAAGTGTATATACTTACCTCTTAATAATAAACTCTATCTATGATGATGTTTC 730  
Db 868 AGCTTATATTCGAAAAATGTCATGTTTATCTTAATCCTTACCTGAGGAAAGTATA 809

QY 731 TTTGTTTCATCATCATCATCATGATGATGATGATGATGATGATGATGATGATGATGAT 790  
Db 808 GGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749

QY 791 AAAAAA 808  
Db 748 AAAAAA 731

RESULT 14  
US-09-969-347-216/c  
; Sequence 216, Application US/09969347  
; Patent No. US20020115085A1  
; GENERAL INFORMATION:  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
; FILE REFERENCE: 689290-69  
; CURRENT APPLICATION NUMBER: US/09/969,347  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: US/60/237,598



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 10:41:08 ; Search time 2183 Seconds

(without alignments)  
6001.900 Million cell updates/sec

Title: US-09-822-080B-1  
Perfect score: 809  
Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum Match 100%  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.8	72.5	685	17	BH711815
2	531.4	65.7	830	17	BH606894
3	293	36.2	660	17	BH456221
4	228	28.2	726	17	BH457950
5	222.4	27.5	604	10	AV832159
6	205.2	25.4	773	17	BH711232

c	7	205	25.3	742	17	BH499547	BH499547
c	8	195.2	24.1	651	17	BH420085	BH420085
c	9	184.8	22.8	632	17	BH604620	BH604620
c	10	183.8	22.7	374	17	BH514321	BOHJX01TR
c	11	183.8	22.7	752	17	BH544042	BOG2E25TR
c	12	170.8	21.1	742	17	BH491802	BOH1M14TF
c	13	162.6	20.1	566	14	N38333	19560 Lambd
c	14	158.8	19.6	566	17	BH78808	BH678808
c	15	154.6	19.1	586	10	AV345497	BOMNB30TF
c	16	151.2	18.7	556	9	AI998304	AV545497
c	17	131	16.2	441	14	R30257	701545163
c	18	128.8	15.9	694	17	BH511873	R30257 12862 Lambd
c	19	122.2	15.1	788	17	BH496475	BH511873
c	20	98.6	12.2	799	17	BH696173	BOGVE05TF
c	21	95.2	11.8	828	17	BH568622	BOGK60TF
c	22	93	11.5	381	10	AV823955	BOMN43TF
c	23	84	10.4	775	17	BH511814	BOG2E25TR
c	24	78.8	9.7	422	10	AV819408	BOHKO78TF
c	25	76.6	9.5	418	10	AV788475	AV819408
c	26	69.6	8.6	718	17	BH516975	AV788475
c	27	68.6	8.5	721	17	BH717953	BH516975
c	28	64.8	8.0	482	10	AW831672	BOGXX15TR
c	29	64	7.9	533	12	BG363083	BH717953
c	30	63	7.8	742	10	BE659870	BOMMV92TR
c	31	62	7.7	528	10	AW395947	AW831672
c	32	62	7.7	542	12	BG882612	sm05f04.y
c	33	62	7.7	550	14	BQ080503	BE659870
c	34	62	7.7	554	14	BQ080567	GM700011A
c	35	62	7.7	571	13	BI701280	AW395947
c	36	62	7.7	581	14	BQ785236	sh07g04.y
c	37	62	7.7	584	14	BQ473539	sa94h05
c	38	62	7.7	593	10	AW760364	BQ785236
c	39	62	7.7	603	12	BF009406	sa94h05
c	40	61.8	7.6	831	17	BH477472	AW760364
c	41	61	7.5	610	10	AW832055	ss78d02.y
c	42	60.6	7.5	540	10	AW201420	BH477472
c	43	60.2	7.4	411	12	BG045717	BOHMX70TR
c	44	60.2	7.4	446	10	BE190618	AW832055
c	45	60.2	7.4	562	14	BQ473476	AW201420

## ALIGNMENTS

RESULT 1  
LOCUS BH711815  
DEFINITION BOHYO04TR BO\_2\_3\_KB Brassica oleracea genomic clone BOHYO04, DNA  
685 bp DNA linear GSS 20-FEB-2002  
ACCESSION BH711815  
VERSION BH711815.1  
KEYWORDS GT:18802293  
SOURCE GSS.  
ORGANISM Brassica oleracea.  
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 685)  
TOWN, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHYO04TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1. .685

/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHFO04"  
/clone\_lib="BO\_2\_3\_KB"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"  
BASE COUNT 176 a 164 c 171 g 174 t  
ORIGIN

Query Match 72.5%; Score 586.8; DB 17; Length 685;  
Best Local Similarity 92.9%; Pred. No. 6.6e-89;  
Matches 637; Conservative 0; Mismatches 47; Indels 2; Gaps 2;

```
QY 8 CCTATGTTTACTTCTCTTCCCTTACCACCTGTTTGGCGGCGACCCCAACAGCTGGA 67
D 1 CCTATGTTTACTTCTCTTCCCTTAAACAGCTGTTTGGCGGCGACCCCAACAGCAGA 60
QY 68 CCAGTTCTCGACACTGATGTTGATATCATATTCCAGCGCAGTTACTACGTTCTCCCTCC 127
D 61 CCAGTTCTCGACATCGATGTTGATATCATATTCCAGCGCAGTTACTACGTTATCCCGTC 120
QY 128 ATCTGGGGCCCTACAGGTGGCGCCCTAACTCTGCTCTCCGCTGGTGGCAACCAAGTGTCC 187
D 121 ATCCGGGGCCCTAAGGTGGCGCCCTAACTCTCACCACCGCAACGGCAACCAAGTGTCC 180
QY 188 CTCTTTATCGACAGAGCGTTTCAGAGTCAACAGGCGCATCCCGTGAAATCTCAAC 247
D 181 CTCTTTATCGGACAGAGCGTTTCAGAGTTCGAAAGGGGCAATCCCGTGAAATCTCAAC 240
QY 248 TGGAGGTCCAGAGTTGGGTTGCTCCCGCAAGAGAGAACTCAACATCAAGATGATGTC 307
D 241 TGGAGGTCTAGAGTTGGGTTGCTCCCGCAATCCGAGAACTCAACATCAAGATGATGTC 300
QY 308 GAACCTAGATCTGCGCTAGTCTAGTCTATTTGTTGGTCACTCCAGCCCGGCTCCCTGG 367
D 301 GAACCTAGTCTGCGCTAGTCTAGTCTATTTGTTGGTCACTCCAGCCCGGCTCCCTGG 360
QY 368 AGTCTGTTTATAGCGGCTGGTCTAGCCAGAGAGCTGGAGGAGAACTCGTCGAGG 427
D 361 AGTCTGTTTATAGCGGCTGGTCTAGCCAGAGAGCTGGAGGAGAACTCGTCGAGG 420
QY 428 AGTTTCTTCAGATCAAGAAACTGAAGCAAACTTAAGCGTTTACAGTTTGTATTTCTGT 487
D 421 AGTTTCTTCAGATCAAGAAACTGAAGCAAACTTAAGCGTTTACAGTTTGTATTTCTGT 480
QY 488 AGTCAGGTAAACATTCATCTGATGTCGTTAAACAGAGAGGTGGCGTTGCGGGTTTG 547
D 481 AGACAGGTAAACATTCATCTGATGTCGTTAAACAGAGAGGTGGCGTTGCGGGTTTG 540
QY 548 GTTTTAGGCTCTACGCCACCATTCGTTACCCCATTCGAGGTTTGTTCGTGAAGCTACT 607
D 541 GTTTTAGGCTCTAGGCCACCATTCGTTACCCCATTCGAGGTTTGTTCGTGAAGCTACT 600
QY 608 GGGACAGACTTCACTCAAGACTATGTTCTATTAATCTGAGAGAAATTAAGACCACTTAA 667
D 601 GGGTCAGAAACTTCACTCAAGACTATGTTCTATTAATCTGAGAGAAATTAAGACCACTTAA 659
QY 668 TAAAGAGGATAAGTGTATTAACATTAC 693
D 660 T-AAGGGGATAAGAGTTTATAACTTAC 684
```

RESULT 2  
LOCUS BH606894 830 bp DNA linear GSS 15-DEC-2001  
DEFINITION BOGFM83TR BOGF Brassica oleracea genomic clone BOGFM83, DNA  
sequence.  
ACCESSION BH606894  
VERSION BH606894.1 GI:17859340  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 830)  
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other\_GSSs: BOGFM83TF  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.  
Location/Qualifiers  
1..830  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGFM83"  
/clone\_lib="BOGF"  
/note="vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"

FEATURES  
source

BASE COUNT 213 a 185 c 193 g 239 t  
ORIGIN

Query Match 65.7%; Score 531.4; DB 17; Length 830;  
Best Local Similarity 93.1%; Pred. No. 1e-79;  
Matches 567; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

```
QY 1 GATGAATCCTATGTTTACTTCTCTTCCCTTACCACTTATTCACCGGCGACCGAAA 60
D 621 GATGAATCCTATGTTTACTTCTTCTTCCCTTAACTGCTTAACTGCTGTTTAGCGGCGACCGAAA 562
QY 61 CGCTGGACCACTTCTCGACACTGATGATATCATATTCGACGGCAGTACTACGTTCT 120
D 561 CGCAGGACCACTTCTCGACACTGATGATGATATCATATTCACCGCAGTACTACGTTAT 502
QY 121 CCCCTCATCTGGGGCCCTACAGGTGGGGCCCTAACTCTCGTCTCCCTCGTGGCAACCA 180
D 501 CCCCTCATCTGGGGCCCTGAAGTGGGGCTTAACTCTACCAACCGGCAACCA 442
QY 181 GTGTCCTCTTATTCGGACAGGAGCGTTTCAGAGTCAACAGGGGCATTCCTCGTGAAT 240
D 441 GTGTCCTCTTATTCGGACAGGAGCGTTTCAGAGTCAACAGGGGCATTCCTCGTGAAT 382
QY 241 CTCAACTGGAGGTCAGAGTTGGGTTCCGTCGCCGAAGAAGAACCTCAACATCAAGAT 300
D 381 CTCAACTGGAGGTCAGAGTTGGGTTCCGTCGCCGAATCCGAGAACCTCAACATCAAGAT 322
QY 301 GGATGTCGAACCTAGCATCTGGCTCAGTCAGCTATTGGTGGGTCACTCCAGCCCCCAG 360
D 321 GGATGTCGAACCTAGCTTCTGCGCTCAGTCACTATTGGTGGGTCACTAGAGCCCCCAG 262
QY 361 TCCCTGGAGGTCGTTGTTTCATAGCGGCTGGTCCCTAAGCCAGAAGCTGGAGAGAAGATTC 420
D 261 TCCCTGGAGATCGGGGTTTCATAGCGGTTGGTCCCTAAGCCAGAAGCTGGAGAGAAGATTC 202
QY 421 GTCGAGGAGTTTCTTCCAGATCAAGAAACTGAAGCCAAACTTAACGCTTACAAGTTTGT 480
D 201 GTCGAGGAGTTTCTTCCAGATCAAGAAACTGAAGCCAAACTTAACGCTTACAAGTTTGT 142
QY 481 ATTCTGTAGTCAGGCTAACGATTGTCATCGTATCGTATCGTATCGTATCGTATCGTATCG 540
D 141 ATTCTGTAGAGCGGTAACTGATTCGATCTCGTATCGTATCGTATCGTATCGTATCGTATCG 82
QY 541 GGGTTTGGTTTTAGGCTCTACGCCACCATTCGCTACCCCATTCGAGGTTGTGTTCTGTA 600
D 81 GGGTTTGGTTTTAGGCTCTAGGCCACCATTCGCTACCCCATTCGAGGTTGTGTTCTG -GAA 23
```



QY 601 AGCTACTGG 609  
 Db 22 AGCTACTGG 14

## RESULT 3

BH456221/c  
 LOCUS BOHRQ52TR BOHR Brassica oleracea genomic clone BOHRQ52, DNA  
 DEFINITION sequence. GSS 12-DEC-2001  
 ACCESSION BH456221  
 VERSION BH456221.1 GI:17641932  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 660)  
 Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other GSSs: BOHRQ52TR  
 Contact: Chris Town

## TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: Tr

Class: sheared ends.

Location/Qualifiers

1..660

/organism="Brassica oleracea"

/strain="Tol000DH3"

/db\_xref="taxon:3712"

/clone="BOHRQ52"

/clone\_lib="BOHR"

/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHS01 using BstXI linkers"

193 a 163 c 145 g 159 t

## BASE COUNT

## ORIGIN

Query Match 36.2%; Score 293; DB 17; Length 660;  
 Best Local Similarity 69.9%; Pred. No. 7.9e-40;  
 Matches 473; Conservative 0; Mismatches 185; Indels 19; Gaps 5;

QY 77 GACACTGATGATATCATATTCGACGGGAGTTACTACGTCTCCCGCTCATCTGGGC 136

Db 660 GACAGTGGTGGATGATACATAACCAACGGCAGTACTTTTCTCCCGCTGGCATCGCGGC 601

QY 137 CCTACAGTGGCGGCTAACCTCTGCTCTCCCGTCTGGCAACAGTGTCCCTCTTTATC 196

Db 600 GCTATTGGTGGCGGCTGAATCTCTCCCGCATCTTGGCTAAATGTCCCATATATT 541

QY 197 GCACAGGAGCGTTGAGAGTCAACAGGGGATTCCTCGTAAATTTCTCAACTGGAGGTC 256

Db 540 GGGCAGGAAGATCCAAAGGTCAACAGGGGCTTCCAGTTAAATTTCTCAACTGGCATCT 481

QY 257 AGAGTTGGTTCGTCCTCCGAGAGAGAACCTCAACATCAAGATGGATGTGCAACCTACG 316

Db 480 AAGTGTGGGTTCTGCTCTGATGAGAGGCTCAACATCGAGATGGATGCAAAAGTACG 421

QY 317 ATCTCGCTCAGTCAGTCTATTGTGGTGTCTACTCCAGCCCGCCAGTCCCTGGAGTCTTG 376

Db 420 ATTTGTGTCACAGCAACTATTGTGGGGCATTTGGGACCATTCATCGGTACATCGCTT 361

QY 377 TCATAGCGGCTGTCTCAAGCAGAGACTGGAGAGAGATTCTGTCGAGAGTTTCTTC 436

Db 360 TCCATAAAGGTTGGTCTCAAGCCAGGTGCT--GGAAAGATTCGATCAGCGCTTCTTT 304

QY 437 CAGATCAAGAAACTGAAGCAAACTTAACGCTTACAAGTTTGTATTCTCTAGTAGGGT 496

Db 303 CAGATCAAGAAACTGAAG---ATGTTGGCGTTTACAATATTTCGAGTTCTCTGTCGT 247

QY 497 AACGATTGCATCGATCGTGTAAACAGAGAGAGTGGCGTTCGGGGTTTGGTTTATAGGC 556

Db 246 AACACTTGCAGAGACTGTTGGTTACTTGTGGTGGTGGCGTTCGACGTTTGGTTGCAAGG 187

QY 557 TCTAGCCACCATTCGCTACCCCATTCGAGGTTGTGTTCTGAAAGCTACTGGGAC-AGA 615

Db 186 TTT-----GCATACGCTAAGCCATTCCCGGTTCCGCTCGTGAAGCTACTGGGACTTCG 133

QY 616 CACTTCATCAAGACTATGCTATTATCTCAGAGAAATTAAGACCACCTTAAATAAGAGG 675

Db 132 TACTTCGTCGAAGACTATGCTATTATCTCAGAGAAATTAAGACTACAAACTAAG--- 76

QY 676 ATAGTCTTAACTTACCTCTAATAATAAACTCTATCTATGATGATGATGATGATGATG 735

Db 75 ---AGAGTTATACTACGACTGATAATAATACTCATCTGCTGCTGCTGCTGCTGCTG 19

QY 736 TCATCGATCATCATCAT 752

Db 18 TCATCGACCACTCTCTT 2

## RESULT 4

BH457950/c

LOCUS BOHRQ52TR BOHA Brassica oleracea genomic clone BOHRQ5, DNA

DEFINITION sequence.

ACCESSION BH457950

VERSION BH457950.1 GI:17643661

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 726)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOHRQ5TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: Tr

Class: sheared ends.

Location/Qualifiers

1..726

/organism="Brassica oleracea"

/strain="Tol000DH3"

/db\_xref="taxon:3712"

/clone="BOHRQ5"

/clone\_lib="BOHA"

/note="Vector: pHS01; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHS01 using BstXI linkers"

207 a 145 c 160 g 214 t

## BASE COUNT

## ORIGIN

Query Match 28.2%; Score 228; DB 17; Length 726;

Best Local Similarity 73.8%; Pred. No. 5.4e-29;

Matches 304; Conservative 0; Mismatches 105; Indels 3; Gaps 1;

QY 1 GATGAATCTATGTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 60

Db 443 GATGAACCTATGTTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 384

QY 61 CGCTGG---ACAGTTCTCGACACTGATGATGATGATGATGATGATGATGATGATGATG 117

Db 383 TGGCGGCAACAGTTCTCGACAGTGATGGTGATATCATAAACAGCGAGTTACTTTGT 324  
 Qy 118 TCTCCCCCTCANTCTGGGCGCCCTACAGGTGGCGGCTTAACCTCTCGTCTCCGCTCGTGGCAA 177  
 Db 323 TCTCCCTGCATCTGGTGGCGCTATTGGTGGCGGCTGAATCTCTGGCGGATCCCTTGGCT 264  
 Qy 178 CAGTGTCCCTCTTTATCGGACAGAGGCTTCAGAGGTCAACAGGGGCTATCCCGTGAA 237  
 Db 263 AATGTCCCTATTATTTGGCAGAGAGTTCAAGGTGAACAGGGGCTTTCAGTTAA 204  
 Qy 238 ATTCTCAAACTGGAGTGCAGAGTTGGGTTCGTCGCCGGAAGAGAGACCTCAACATCA 297  
 Db 203 ATTCTCAAACTGGCAGTCTAAAGTTGGGTTCGTTCTGAATCAGAGAGCTCAACATCGA 144  
 Qy 298 GATGGATTCGAACTACGATTCGCTCAGTCAGTTATTTGGTGGTCACTCCAGCCCC 357  
 Db 143 GATGGATTCGAACTACGATTTGTCAGGCAACCTATTGGTGGGCAATTCGGGAGCA 84  
 Qy 358 CAGTCCCTGGAGTCTGTTTCATAGCGGCTGGTCTCAAGCCAGAGCTGGA 409  
 Db 83 TCCATGGTTACATCCCTTCCCATAAAGTTGGTCTCAAGCCAGGTGGCGAA 32  
  
 RESULT 5  
 AV832159 604 bp mRNA linear EST 01-APR-2002  
 LOCUS AV832159 RAF11 Arabidopsis thaliana cDNA clone RAF11-04-122 5',  
 DEFINITION mRNA sequence.  
 ACCESSION AV832159  
 VERSION AV832159.1 GI:19874219  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 604)  
 Seki,M., Narusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,  
 Oono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,  
 Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.  
 and Shinozaki,K.  
 Large scale analysis of Arabidopsis full-length cDNA (2002b)  
 Unpublished (2002)  
 Contact: Motoaki Seki  
 Plant Functional Genomics Research Group  
 RIKEN Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: 81-298-36-4359  
 Fax: 81-298-36-9060  
 Email: mseki@tc.riken.go.jp  
 An Arabidopsis full-length cDNA library was constructed essentially  
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI  
 and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et  
 al., submitted for publication) digested with BamHI and SalI. This  
 clone is in a modified phagescript vector. Please visit our web  
 site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further  
 details.  
 FEATURES  
 Location/Qualifiers  
 1..604  
 /organism="Arabidopsis thaliana"  
 /db\_xref="taxon:3702"  
 /clone="RAF11-04-122"  
 /dev\_stage="plants at various developmental stages from  
 germination to mature seeds"  
 /lab\_host="DH10b"  
 /note="Site\_1: BamHI; Site\_2: SalI; subjected to various  
 treatments (dehydration, cold, high salt, ABA, heat and UV  
 ). Dark-grown plants"  
 BASE COUNT 153 a 142 c 156 g 152 t 1 others  
 ORIGIN  
 Query Match 27.5%; Score 222.4; DB 10; Length 604;

Best Local Similarity 69.5%; Pred. No. 5.2e-28;  
 Matches 349; Conservative 0; Mismatches 141; Indels 12; Gaps 3;  
 Qy 2 ATGAATCCTATGTTTACTTCTCTTCTTCCCTTTACCACTGTTTGGCGCGACGCGAAAC 61  
 Db 50 ATGAATCCTAAGTTTACTTGGTTCTTAAACCGCGTTTGGCCTCAACAGCATAT 109  
 Qy 62 GCTGGACACAGTTTCGACACTGATGTGATATCATATATTCGACGCGACTTACTACGTTCTC 121  
 Db 110 GGTG---CGGTTGTAGACATCGATGAAACGCCATGTTCCACGAAAGTTACTAGTTCTC 166  
 Qy 122 CCCCTCATCTGGGCGCCTACAGGTGGCGGCTAACTCTCTCTCTCCCTGCTGCGAACCCAG 181  
 Db 167 CCTGTCATCCGTGGCC---GAGCGGAGGCTGACTCTAGCAGGCGCGGTGGCGAGCCA 223  
 Qy 182 TGTCCCTCTTTATCGACAGGCGCTTCAGAGTCAACAGGGGCATTCCTCGTGAAATTC 241  
 Db 224 TGTCTTACGATATCGTGCAGGAATCTTCAGAAGTTTGATGAGGCGATTCCTCGTAAATTC 283  
 Qy 242 TCAAACTGGAGTCCAGAGTTGGTTCGTCGCCGGAAGAGAACCTCAACATCAAGATG 301  
 Db 284 TCAAACTGGAGGCTTAAGTTGCGTTCGTTCCCGAATCACAGAACCTCAACATCGAAACA 343  
 Qy 302 GATGTGAACTAGATCTCGCTCAGTCAGCTTATTTGGTGGTCACTCCAGCCCCAGT 361  
 Db 344 GACGTGGAGCGACGATCTGCATCCAGTCAACCTACTTGGCGGTTCGGTGGAGTTGACCCAC 403  
 Qy 362 CCCTGGAGTCTGTTTCATAGCGGCTGGTCTTAAGCCAGAGCTGGAGGAGAGATTCG 421  
 Db 404 GAGAGGAGGAGTACTTCTGCTGCTGCTCAAGGCGAGGAGGTTCCGACAGATTCG 463  
 Qy 422 TCGAGGAGTTTCTTCCAGATCAAGAAACTGAAGCCAACTTAACGTTACAAATTTGTA 481  
 Db 464 TTGAAGATTTTCTCAAGATCGAGAAATCTGGAG-----AGGATGCTTACAAGTTTGTG 517  
 Qy 482 TTCGTAGTGGGTAAAGATT 503  
 Db 518 TTCGTCTCGGAGTTCGGACT 539  
  
 RESULT 6  
 BH711232 773 bp DNA linear GSS 20-FEB-2002  
 LOCUS BH711232 BO\_2\_3\_KB Brassica oleracea genomic clone BOHT192, DNA  
 DEFINITION sequence.  
 ACCESSION BH711232  
 VERSION BH711232.1 GI:18801018  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 773)  
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other\_GSSs: BOHT192Tr  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: Tg  
 Class: sheared ends  
 Location/Qualifiers  
 1..773  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHT192"  
 FEATURES  
 source

```

/clone_lib="BO_2_3_KB"
/Note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      204 a 151 c 191 g 227 t
ORIGIN

Query Match      25.4%; Score 205.2; DB 17; Length 773;
Best Local Similarity 70.6%; Pred. No. 3.3e-25;
Matches 370; Conservative 0; Mismatches 123; Indels 31; Gaps 6;

QY 166 CCGTCGTGGCCACACGATGTCCTTTATCGGACAGGAGCGTTGAGAGGTCAACAGGGG 225
Db 20 CAGTGTGTGAGGATGTCCTTTATCGGACAGGAGCGTTGAGAGGTCAACAGGGG 79

QY 226 CATTCCTGCGAATCTCAAACTGGAGGTCAGAGTTGGTTCGTCGCCGAGAGAGAGAA 285
Db 80 CATTCCTGCGAATCTCAAACTGGAGGTCAGAGTTGGTTCGTCGCCGAGAGAGAA 139

QY 286 CCTCAACATCAGATGGATGTCGACCTACGATCTGCGCTCAGTCAGCTTATTTGGTGGT 345
Db 140 CCTCAACATGAGATGGAGTCAAGCTACGCTCTGCGTCAAGTCAAGTCAAGTCAAGT 199

QY 346 CACTCCAGCCCCAGTCCCTGAGGTCGTTGTTTCATAGCGGCTGTCCTCAAGCCAGAGC 405
Db 200 CTCGGCGACAGGGCAGTAAGCTGGCTG-----GAGGCTGTCGGAAGCCAGATGA 253

QY 406 T---GGAGGAGAGATTCCTGAGGAGTTTCTTCCAGATCAAGAAAC---TGAAGCCAA 459
Db 254 TCAGCTGACAGAGATTCCTGAGGAGTTTCTTCCAGATCAAGAAAGCCGATGGAGGAT 313

QY 460 ACTTAACGTTTACAAGTTTCTGATGAGGTTAAGATTTCATCGATCGATGCGTAA 519
Db 314 ACTTAGGGTTTACAATATGCGTTTGTCTTAAAGATGAGGATTCATCGATGTCAGTGT 373

QY 520 AAACGAGGAGG---TGGGTTTCGGGTTGGTTTGGTTTGGGTTTGGGTTTGGGTTTGG 419
Db 374 ATATTGGAAAGAACTACCGTTGGGTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTTT 419

QY 577 CCATTCGAGGTTGTTGCTG---GAAAGCTACTGGGACAGACACTTCATCCAGACTATGT 635
Db 420 -CCAATTCGAGGTTGATGTTGCTGTTAAAGCTACTAGGACAGAGACTTTGTCAGACTATGT 478

QY 636 CTATTATCTGAGAGAAATTAAGACCCTTAATTAAGAGGATAA 679
Db 479 CTATTATCTGAGAGATAAGTGTATACTCAAGACTAATAATAA 522

RESULT 7
BH499547/c
LOCUS      BH499547
DEFINITION BOGVA20TR BOGV Brassica oleracea genomic clone BOGVA20, DNA
sequence.
ACCESSION  BH499547
VERSION    1
KEYWORDS   GSS.
SOURCE     Brassica oleracea.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Other_GSSs: BOGVA20TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR

```

```

Class: sheared ends.
Location/Qualifiers
1. .742
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOGVA20"
/Note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      245 a 155 c 131 g 211 t
ORIGIN

Query Match      25.3%; Score 205; DB 17; Length 742;
Best Local Similarity 68.2%; Pred. No. 3.7e-25;
Matches 406; Conservative 0; Mismatches 120; Indels 69; Gaps 6;

QY 189 TCTTTATTCGGACAGAGCGTTTCAGAGGTCAACAGGGGCATTCCTCGTGAATTTCTCAACT 248
Db 742 TCTATATTCGGACAGATATTCAGAGGTCAACAGGGGCATTCCTCGTGAATTTCTCAGACT 683

QY 249 GGAGGTCCAGAGTTGGTTCCTCCCGAAGAAGAGAACCTCAACATCAAGATGGATGTCG 308
Db 682 GGAGGACTAAAGTTGCGTTCCTCCTGAATCAACGAACCTCAACATAGATGGAGCTCA 623

QY 309 AACCTACGATCTCGCTCAGTCAGCTTATTTGGTGGGTCACTCCAGCCCCCAGTCCCTGGA 368
Db 622 AAGCCAGCATATGTTGTCAGTCAACCTTATTTGGTATGTACCTGCACCCGACATGTTGTTG 563

QY 369 GGTCTGTTGTCATAGCGCTGGTCTAAGCAGAGCTGAGGAGAGATTCCTCGAGGA 428
Db 562 AGCGGCGTTCATAGCGCTGGTCTTAACCA-----TCGAATG 524

QY 429 GTTTTCTCCAGATCAAGAAACTCAAGCCAAACTTAACGCTTACAAGTTTGTATTTCTGTA 488
Db 523 ATTTCTCCAGATCAAGAAATTTGAAGATTTCTATTTGAGGTTTACAAGATTTGTTTGT 464

QY 489 GTGAGGTTAAGATTCGATCGATGTCGGTAAACACAGG---AAGTGGGTTTCGGGGTT 545
Db 463 TTAACGTTGAAGATTCGGTTCGATGTCGGGATATTTTGGACAAACATGCGGTTAGCGTT 404

QY 546 TGGTTTATGCTCTACGCCACCATTCGCTACCCCATTCGAGGTTGTTGTCGTAAGAGCTA 605
Db 403 TGGCTTATGCTCTACGCC-----TTCGAGGTTGTTGTCGTAAGAGCTA 359

QY 606 CTGGGACAGACTTTCATCCAGACTATGTCATTATTCGAGAGAAATTAAGACCACTT 665
Db 358 GTGAGACAAAGACTTCGTCACAGCTGTTATGCTATTAT----- 319

QY 666 AATAAGAGGATAAGTGTATATACTTACCTCTAATAATAAACTCTATCTATGATGATG 725
Db 318 ----AGAGGATAAGATTTATAAGCTACG---AATAATAAACTCTATCTGTTTCATG 266

QY 726 TTTTCTTTGTTTCATCATCATCATCATGATGATGATGATGATGATGATGATGATGATG 780
Db 265 TTTTCTTTGTTTCATCATCATCATGATGATGATGATGATGATGATGATGATGATGATG 214

RESULT 8
BH420085/c
LOCUS      BH420085
DEFINITION BOGUK16TR BOGU Brassica oleracea genomic clone BOGUK16, DNA
sequence.
ACCESSION  BH420085
VERSION    1
KEYWORDS   GSS.
SOURCE     Brassica oleracea.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
TOWN,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

```

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOGUK16TF  
Contact: Chris Town  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

#### FEATURES

source  
1. .651  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGUK16"  
/clone\_lib="BOGU"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 185 a 126 c 136 g 204 t

ORIGIN

Query Match 24.1%; Score 195.2; DB 17; Length 651;  
Best Local Similarity 77.8%; Pred. No. 1.7e-23;  
Matches 249; Conservative 0; Mismatches 68; Indels 3; Gaps 1;  
Qy 1 GATGAATCCATGTTTACTTCCTTCTTGCTTTTACACATGTTTGGCGGCGACCGCAA 60  
Db 320 GATGAACCTATGTTTACTTCCTTATTCCTTAAACCGTTGTTTGGCCACCACAGCAA 261  
Qy 61 CGCTGG---ACCAAGTCTCCACACTGATGTGATATCATATTTCGACGCGAGTTACTACGT 117  
Db 260 TGGCGGGAACCAAGTCTTCGACAGTGTGATGATATCAATCAACCAAGCGAGTTACTTGT 201  
Qy 118 TCTCCCTCATCTGGGCGCCTACAGTGGCGCCTAACTCTCGTCTCCGCTGGTGGCAA 177  
Db 200 TCTCCCTGCCATCTGCGGCGCTATTGTGGCGGCGCTGAATCTCTCGCGCGATCCTTGGCT 141  
Qy 178 CCAGTGTCCCTCTTTATCGACAGAGCGCTTCAGAGGTCAACAGGGGCAATCCCGTGAA 237  
Db 140 AAAATGTCCCTATTATATGGGAGGAAGGTTCAAAGGTTGAACAGGGGCTTTCCAGATTAA 81  
Qy 238 ATTCTCAAACTGGAGGTTCAGAGTTGGTTTCGTCCTCCCGAAGAGAAACCTCAACATCAA 297  
Db 80 ATTCTCAAACTGGAGGTTCAGAGTTGGTTTCGTTCTCTGATCAGAGAGGCTCAACATCGA 21  
Qy 298 GATGGATGTCGAACCTACGA 317  
Db 20 GATGGATGGCAAGCTACGA 1

RESULT 9  
BH604620 632 bp DNA linear GSS 15-DEC-2001  
LOCUS BOHF021TR BOHF Brassica oleracea genomic clone BOHF021, DNA  
DEFINITION sequence.

ACCESSION BH604620  
VERSION BH604620.1 GI:17857066  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 632)  
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
REFERENCE  
AUTHORS Whole genome shotgun sequencing of Brassica oleracea  
TITLE Unpublished (2001)  
JOURNAL Other\_GSSs: BOHF021TF  
COMMENT Contact: Chris Town  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

#### FEATURES

source  
1. .632  
/organism="Brassica oleracea"  
/strain="T01000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHF021"  
/clone\_lib="BOHF"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 157 a 144 c 155 g 176 t

ORIGIN

Query Match 22.8%; Score 184.8; DB 17; Length 632;  
Best Local Similarity 61.0%; Pred. No. 9.6e-22;  
Matches 382; Conservative 0; Mismatches 217; Indels 27; Gaps 4;  
Qy 2 ATGAATCCTATGTTTACTTCTTCTTCCCTTTACCACGTGTTTGGCGGCGACCGCAAAC 61  
Db 34 ATGAATCCTATGTTTACTTCTTCTTCCATCAACCCCTAGTTTGGTTACAAACACATAT 93  
Qy 62 GCTGACACAGTTCGACACATGATGGTGATATATTCGACGCGAGTTACTACTGTTCTC 121  
Db 94 GGTG---CTGTCTTGACACCGCGGACATATTTCCGTCGCGAGTTACTATGTTCTC 150  
Qy 122 CCCCTCATCTGGGCGCCTACAGTGGCGGCTAACTCTGCTCCGCTCGTGGCAACAG 181  
Db 151 CCCGTCGTCGCGGCGC---GAGGAGGCGGCTTGAATCTAGGCGCGCGTGGGAGCTA 207  
Qy 182 TGTCCCTCTTTATCGACAGGAGGCTTCAGAGTCAACAGGGGCATTCCTCGTGAATTC 241  
Db 208 TGTCTTACGACATCGTGCAGAAATCATCTGAATCTGACGAGGATTTCCCGTTAAATTC 267  
Qy 242 TCAAACTGGAGTCCAGAGTTGGGTTCTGCCCGAAGAGAGAACCTCAACATCAAGATG 301  
Db 268 TCGAATCGAGACCTAGAGTTGCGTTTCTCTGAGTCACAAGACCTTAACATCAAGACG 327  
Qy 302 GATGTCGAACCTACATCTGCGCTCAGTCAGTTATTTGGTGGTCACCTCCAGCCCCCAGT 361  
Db 328 GACGTTGAAGCTACGATATGCTTCCAGTCAACATCTGAGAGTGGTGGTGGTGGACG 387  
Qy 362 CCCCTGAGTCTGCTTCTATAGCGGCTGCTCTAAGCCAGAGCTGGAGGAGAGATTTCG 421  
Db 388 GAGAGGACGAGTATTTCTGTGGTGGTGGTCTGTC-----TACAGAGCAGCTCA 432  
Qy 422 TCGAGGAGTTTCTCCAGATCAAGAAACTGAAGCCAACTTAAGCTTTACAAGTTTGTGA 481  
Db 433 CCCAACAGTTTCTCCAGATCGAAAAATCTGGAG-----ATGATGCTTACAAGTTTGTG 486  
Qy 482 TTCTGTAGTACGGTTAAGATTCGATCGATCTCGTGAAGAGAGAGAGGTCGCTTCGG 541  
Db 487 TTCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546  
Qy 542 GGTGTTGGTTTATAGGCTCTACGCCACATTCGCTACCCCATTCGAGGTTGCTGCTCGTGA 601  
Db 547 GAATAGGCGTTGCGGCTTTGCTTTAAGTCTGAGCCGCTTCTGTTATGTTTCAAGAA 606  
Qy 602 GTACTGGGACAGACACTTTCCTCA 627  
Db 607 GCTAATGTTACCGAGATATCATCCGA 632

RESULT 10  
BH514321/c  
LOCUS BOHJX01TR BOHJ Brassica oleracea genomic clone BOHJX01, DNA  
DEFINITION sequence.

BH514321 374 bp DNA linear GSS 13-DEC-2001  
BOHJX01TR BOHJ Brassica oleracea genomic clone BOHJX01, DNA

ACCESSION BH514321  
 VERSION BH514321.1 GI:17722411  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 374)  
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOHJX01TF  
 Contact: Chris Town  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.  
 FEATURES  
 source  
 1..374  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHJX01"  
 /clone\_lib="BOHJ"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"  
 BASE COUNT 95 a 86 c 106 g 87 t  
 ORIGIN  
 Query Match 22.7%; Score 183.8; DB 17; Length 374;  
 Best Local Similarity 76.5%; Pred. No. 2e-21;  
 Matches 267; Conservative 0; Mismatches 72; Indels 10; Gaps 3;  
 QY 5 AATCCTATGTTTACTTCTTCCCTTACCACTGTTTGGCGGACCGCAACCGCT 64  
 DB 374 AATCCTATGTTTACTTCTTCCCTTACCACTGTTTGGCGGACCGCAACCGCT 64  
 QY 65 GG---ACAGTTCGACACTGATGATATATATATTCACGGCAGTACTACGTTCTC 121  
 DB 314 GGCAACAGTTCGACACTGATGATGATATATATATTCACCAATGCGCAGTACTACGTTCTC 121  
 QY 122 CCCCTCATCTGGGGCCCTACAGGTGGCGGCTTAACCTCTGCTCCGTCGTGGCAACCCAG 181  
 DB 254 CCAGTC-----TCCCACTATGAAGGTGGCCCTGACTCTCCCTCCGGTGGCGCAACCCCA 201  
 QY 182 TGTCCCTCTT-TTATCGGACAGGCGGTTTCAGAGGTCAACAGGGGATTCCTCGTGAAT 240  
 DB 200 TGTCCCTCTTATGATCGGACCGGAAATATCAAGGAGGAACAGGGCTTCCCTTAAGATT 141  
 QY 241 CTCAACTGGAGGTCCAGAGTTCGGTTCGTCGCCGGAAGAGAACTCAACATCAAGAT 300  
 DB 140 CTCAACTGGGGTCTGGAGCTAGGTAGTTCGCCGAATCAGAGAACTCAACATCAAGAT 81  
 QY 301 GGATGTGAACCTACGATTCGCTCAGTCAGCTTATTTGGTGGTCACT 349  
 DB 80 GGACCTCCCACTACGATTCGCGTCAGTCATCTATTTGGTGGTCACT 32  
 RESULT 11  
 BH544042/c  
 LOCUS BOHJX01TF BOHJ Brassica oleracea genomic clone BOHJX01, DNA  
 DEFINITION sequence.  
 ACCESSION BH544042  
 VERSION BH544042.1 GI:17795823  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 742)  
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 752)  
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M.  
 TITLE Whole genome shotgun sequencing of Brassica oleracea  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: BOGZE25TF  
 Contact: Chris Town  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.  
 FEATURES  
 source  
 1..752  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGZE25"  
 /clone\_lib="BOGZ"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"  
 BASE COUNT 233 a 131 c 156 g 232 t  
 ORIGIN

Query Match 22.7%; Score 183.8; DB 17; Length 752;  
 Best Local Similarity 76.3%; Pred. No. 1.3e-21;  
 Matches 254; Conservative 0; Mismatches 72; Indels 7; Gaps 2;  
 QY 1 GATGAATCCTATGTTTACTTCTTCTTCCCTTACCACTGTTTGGCGGACCGCAAA 60  
 DB 331 GATGAATCCTATGTTTACTTCTTCTTCCCTTACCACTGTTTGGCGGACCGCAAA 60  
 QY 61 CGC---TGACACAGTCTCGACACTGATGATATATATTCGACGGCAGTACTACGT 117  
 DB 271 CGCGACGCAACAGTGTGTTGACGCTGATGATCTCATATCCGACGCGAGTACTATGC 212  
 QY 118 TCTCCCTCTCATCTGGGGCCCTACAGGTGGCGGCTTAACCTCTCGTCTCCGTCGTTGGCAA 177  
 DB 211 TGTCCCGCTCC---CCGACAATGACGTAGCTGACTCTCGTCTCCGAGGTGGCAA 156  
 QY 178 CCAGTGTCCCTCTTATTCGGACAGGCGTTCAGAGGTCAACAGGGGATTCCTCGTGA 237  
 DB 155 CCAATGTCCTCTTATATCGGACCGGAATATATCAACGAAGAAGAGGCTTGCCTTAA 96  
 QY 238 ATTCTCAAACTGGAGTCCAGAGTTGGGTTGTTCTCCCGAAGAAGAACCTCAACATCA 297  
 DB 95 ATCTCAAACTGGGGTTCATGGGCTGAGTGTGTTCCCGAATCAGAGAACCTGAACATCGA 36  
 QY 298 GATGATGTCAACCTACGATCTGCGCTCAGTC 330  
 DB 35 GATGAACGTCCCATCTACGATCTGCGGTCACTC 3

RESULT 12  
 BH491802  
 LOCUS BOHIM14TF BOHI Brassica oleracea genomic clone BOHIM14, DNA  
 DEFINITION sequence.  
 ACCESSION BH491802  
 VERSION BH491802.1 GI:17699906  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea.  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 742)  
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHIM14TR  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1. .742  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHIM14"  
/clone\_lib="BOH1"  
/note="Vector: pHS1; Site\_1: BstXI; 2-3 kb sheared  
genomic DNA inserted into pHS1 using BstXI linkers"

BASE COUNT 187 a 156 c 179 g 220 t

Query Match 21.1%; Score 170.8; DB 17; Length 742;  
Best Local Similarity 60.5%; Pred. No. 1.9e-19;  
Matches 346; Conservative 0; Mismatches 202; Indels 24; Gaps 3;

QY 64 TGGACCACTTCTCGACACTGATGGTGATATCATATTCGACGGCAGCTTACTAGTTCCTCC 123  
DB 41 TGGTGCTGTTCTTGACACCGACGCGACATCATTTCCGTTGGCAGTACTATGTTCTCC 100  
QY 124 CCTCATCTGGGGCCCTACAGTGGCGCCCTAACTCTCGTCTCCGTCGGGCCAACAGTG 183  
DB 101 CGTCTCCGGGGCC---GAGGAGGGGGCTTGACTCTAGGCGCGCGGTGGGGAGCTAG 157  
QY 184 TCCCTCTTTATCGACAGGAGCGTTTCAGAGGTCACAGGGGCAATTCCTCGTGAATTC 243  
DB 158 TCCCTAGCATCTGTCAGAAATCATCTGACCTCGAGGAGTATTCCTGTTAAATTC 217  
QY 244 AAACCTGGAGTCCAGAGTTGGGTTGCTCCCGAAGAGAACTCAACATCAAGATGGA 303  
DB 218 GAACCTGGAGACCTAGAGTTGGGTTGTTCTCTGAGTCAACAAGACCTTAACATCAAGAC 277  
QY 304 TGTCGAACCTACGATCTCGCTCAGTCACTATTTGGTGGTCACTCCAGCCCCAGTCC 363  
DB 278 CGTTGAAGCTACGATATGCTTCCAGTCAACATCTGAGAGTGGTGGTTCAGCAGGA 337  
QY 364 CTGGAGGTCTGTTGTTATAGCGCTGCTTAAAGCCAGAGCTGGAGGAGAGATTCGTC 423  
DB 338 GAGGACGAGTATTTCTGTTGGTGGTGTCTACA-----AGACGACTCACC 382  
QY 424 GAGAGTTCTTTCAGATCAAGAAATCAAGCAAACTTAACGCTTACAGTTTCTGATT 483  
DB 383 CAACAGTTCTTTCAGATCGAAAATCTGGAG-----ATGATGCTTACAGTTTGGTT 436  
QY 484 CTGTAGTGAGGTAAACGATCATCGATGTCTGTTAAAGAGGAGGAGGTCGCTTCGGGG 543  
DB 437 CTGTCTCTGATCTGTTGATCTGCTGCTCATGTCAGGAACTGTTGAATATTTGTTGAGCA 496  
QY 544 TTTGGTTTTAGGCTCTACGCCACCATTCGCTTACCCCATTCGAGTTGTTGTTGTTGAAGC 603  
DB 497 AATAGGCGTTCGGGGTTTGGCTTTAAAGTCTTGAGCCGCTTCTTGTGTTATTTCAAGAAAGC 556  
QY 604 TACTGGGACACACTTCATCCAGACTATGT 635  
DB 557 TAATGTTACCGAGATATCATCCGAGCTATGT 588

RESULT 13  
N38333 566 bp mRNA linear EST 05-JAN-1998  
LOCUS N38333  
DEFINITION 19560 Lambda-PRL2 Arabidopsis thaliana cDNA clone 212L20T7, mRNA

sequence.  
N38333  
N38333.1 GI:1159475  
EST.  
thale cress.  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 566)  
Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh  
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel  
,E. and Somerville,C.  
Genes galore: a summary of methods for accessing results from  
large-scale partial sequencing of anonymous Arabidopsis cDNA clones  
Plant Physiol. 106, 1241-1255 (1994)  
95148729  
Contact: Thomas Newman  
MSU-DOE Plant Research Laboratory  
Michigan State University  
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.  
Lansing, MI  
Tel: 517-353-0854  
Fax: 517-353-9188  
Email: 22313tcn@bm.cl.msu.edu  
Seq primer: T7 dye primer.  
Location/Qualifiers  
1. 566  
/organism="Arabidopsis thaliana"  
/strain="var columbia"  
/db\_xref="taxon:3702"  
/clone="212L20T7"  
/clone\_lib="Lambda-PRL2"  
/note="Vector: lambda Zip-Lox; Site\_1: Sal; Site\_2: Not;  
Lambda PRL2 is a cDNA library derived from equal  
quantities of 4 pools of mRNA. The mRNA from equal  
day germinated etiolated seedlings; 2) tissue culture  
grown roots; 3) staged plants half with 24 hour light  
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)  
same plants as 3 but aerial tissue (stems, flowers and  
siliques. The vector is BRL's lambda Zip-Lox. The cDNA  
inserts were directionally cloned with Sal-Not arms using  
oligo dT primed cDNA. 151 g 154 t 22 others

BASE COUNT 120 a 119 c 151 g 154 t

ORIGIN  
Query Match 20.1%; Score 162.6; DB 14; Length 566;  
Best Local Similarity 71.6%; Pred. No. 5.3e-18;  
Matches 242; Conservative 0; Mismatches 90; Indels 6; Gaps 2;

QY 9 CTATGTTTACTTCTCTTTCCTTTTACACACTGTTTGGCGCGACCGCAACGCTGGAC 68  
DB 1 CTAAGTTTACTTGGTTCTTGGCTTAAACCGCGTCTTGGCTTCAACGCGATATGTG--- 57  
QY 69 CAGTTCTCACACTGATGTGATATCATATTCGACGGCAGTTACTACGTTCTCCCTCA 128  
DB 58 CGGTTGACACATGATGGAAGCGCATTTCCAGAAAGTACTACGTTCTCCCTGTCA 117  
QY 129 TCTGGGGCCCTACAGTGGCGGCTAACTCTCTGCTCTCCGCTGTCGCAACAGTGTCC 188  
DB 118 TCCGTGGCC---GAGCGGAGGCGCTGACTCTAGCAGGCGCGGTGGCGAGCCATGTCTT 174  
QY 189 TCTTTATCGACAGGAGCGCTTCAGAGGTCAACAGGGGCAATTCCTGGAATTTCTCAACT 248  
DB 175 ACGATATCTGCGAGGAATCTTCAGAGGCAATTCAGAGGCAATTCCTGGAATTTCTCAACT 234  
QY 249 GGAGGTCAGAGTTGGTTTCGTCGCCGGAAGAGAGACCTCAACATCAAGATGGATGCG 308  
DB 235 GGAGGCTTAAAGTTGCGTTTCCCGCATCACAGAACCTCAACATCGAAGACGCTCG 294  
QY 309 AACCTACGATCTCGCTCAGTCACTAGCTTATTTGGTGGTC 346  
DB 295 GAGCCAGCATCTGCATCCAGTCAACCTCACTACTGGCGGTC 332







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 25, 2003, 18:31:41 ; Search time 2885 Seconds  
(without alignments)  
2158.752 Million cell updates/sec

Title: US-09-822-080B-2  
Perfect score: 1139  
Sequence: 1 MNPWFYLLAFTTTLAATAN.....EVFVRKATGDTTSKTKMSII 214

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4139280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09822080/runat\_22012003\_142641\_4537/app\_query.fasta\_1.391  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09822080.ecgn\_1.1\_1616 @runat\_22012003\_142641\_4537 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMALP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_lm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_or:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pi:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vit:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1139	100.0	783	8	BOU18995	U18995 Brassica ol
2	506.5	44.5	845	8	AY081323	AY081323 Arabidops
3	506.5	44.5	70836	8	AC010536	AC010536 Arabidops
4	496.5	43.6	796	8	AY054566	AY054566 Arabidops
5	246	21.6	858	8	NT066263	U66263 Nicotiana t
6	239.5	21.0	845	8	AF128268	AF128268 Glycine m
7	237	20.8	791	8	LEU70076	U70076 Lycopersico
8	226.5	19.9	1615	8	PSFUC	X82595 P.sativum f
9	226	19.8	967	8	AB075524	AB075524 Oryza sat
10	226	19.8	141545	8	OSJN00221	AL663019 Oryza sat
11	225	19.8	879	8	PSAJ11398	AY011398 Pisum sat
12	222.5	19.5	917	6	I65559	I65559 Sequence 1
13	222.5	19.5	930	8	TC21KD	X56509 T.cacao mRN
14	220.5	19.4	792	8	CAR276263	AJ76263 Cicer ari
15	217.5	19.1	465	8	AF356620	AF356620 Theobroma
16	216.5	18.0	829	8	AKA012840	AJ012840 Avicennia
17	214.5	18.0	916	6	A29723	A29723 S.cerevisia
18	211.5	18.6	465	8	AF356613	AF356613 Theobroma
19	211.5	18.6	465	8	AF356614	AF356614 Theobroma
20	211.5	18.6	3269	8	S45035	S45035 Knil-Kunitz
21	210.5	18.5	465	8	AF356609	AF356609 Theobroma
22	209.5	18.4	633	8	HVBAASI	Z12961 H.vulgare g
23	208.5	18.3	465	8	AF356610	AF356610 Theobroma
24	208.5	18.3	465	8	AF356611	AF356611 Theobroma
25	208.5	18.3	465	8	AF356612	AF356612 Theobroma
26	207	18.2	462	8	AF356607	AF356607 Theobroma
27	207	18.2	2762	8	AF526372	AF526372 Medicago
28	207	18.2	127747	2	AC122730	AC122730 Medicago
29	202.5	17.8	465	8	AF356618	AF356618 Theobroma
30	201.5	17.7	465	8	AF356615	AF356615 Theobroma
31	200.5	17.6	465	8	AF356608	AF356608 Theobroma
32	200.5	17.6	465	8	AF356616	AF356616 Theobroma
33	200.5	17.6	465	8	AF356617	AF356617 Theobroma
34	200.5	17.6	591	6	AB176186	AB176186 Sequence
35	200.5	17.6	591	6	E22901	E22901 Artificial
36	200.5	17.6	603	6	AR176197	AR176197 Sequence
37	200.5	17.6	603	6	AR176198	AR176198 Sequence
38	200.5	17.6	603	6	AR176199	AR176199 Sequence
39	199.5	17.5	465	8	AF356605	AF356605 Herrania
40	199.5	17.5	465	8	AF356606	AF356606 Herrania
41	199.5	17.5	465	8	AF356621	AF356621 Theobroma
42	198.5	17.4	465	8	AF356619	AF356619 Theobroma
43	197.5	17.3	792	8	HVBA51	X16276 Barley mRNA
44	194.5	17.1	480	8	AB023651	AB023651 Solanum m
45	194	17.0	462	8	AF356624	AF356624 Theobroma

ALIGNMENTS

RESULT 1

```

BOU18995          783 bp      mRNA      linear      PLN 04-SEP-1997
LOCUS             Brassica oleracea trypsin inhibitor propeptide mRNA, complete cds.
DEFINITION        U18995
ACCESSION         U18995
VERSION           U18995.1 GI:841207
KEYWORDS          .
SOURCE            Brassica oleracea.
ORGANISM          Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE         1 (bases 1 to 783)
AUTHORS           Williams,D.L., Kain,W.C. and Broadway,R.M.
TITLE             Isolation and characterization of a serine proteinase inhibitor
                  cDNA (Accession No. U18995) from cabbage (PGR97-083)
JOURNAL           Plant Physiol. 114, 747 (1997)
REFERENCE         2 (bases 1 to 783)
AUTHORS           Williams,D.L.
TITLE             Direct Submission
JOURNAL           Submitted (22-DEC-1994) David L. Williams, Entomology, NYSAES
                  Cornell University, Geneva, NY 14456, USA
FEATURES          Location/Qualifiers
                   1..783
                     /organism="Brassica oleracea"
                     /strain="Superpack"
                     /db_xref="taxon:3712"
                     /clone_lib="lambda Zap II"
                     /dev_stage="mature cabbage"
                     2..646
                       /citation=[1]
                       /codon_start=1
                       /product="trypsin inhibitor propeptide"
                       /protein_id="AA868964.1"
                       /db_xref="GI:841208"
                       /translation="WNPMFYLLAFTTVAATNAGPVLDTGDIIFDCSYVPLIIV
                       EPTGGTLVSRGNQCLFIOERSEVNRGIPKFSNMRSGVGVPEEENLNKMDV
                       GPTICQAQYMWVTPAPSPWLSLFAAGPKPEAGGDSRSFFQIKKTEAKLNKMFV
                       FCSEGNDCIDVCKNEEGVRLGLVSTPPFATPEVVFVKATGDTISSTKMSII"
sig_peptide       2..64
BASE COUNT        194 a 186 c 181 g 222 t
ORIGIN
Alignment Scores:
Pred. No.:        3.3e-99      Length:      783
Score:            1139.00      Matches:    214
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:      100.00%      Indels:     0
DB:               8           Gaps:       0

US-09-822-080B-2 (1-214) x BOU18995 (1-783)

Qy 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrThrValLeuAlaAlaThrAlaAsn 20
Db 2 ATGAATCCTATGTTTACTTCTTCTGCTTACCACCTGTTTGGCCGCGCCGCAAC 61

Qy 21 AlaGlyProValLeuAspThrAspGlyAspIlePheAspGlySerTyrValLeu 40
Db 62 GCTGGACAGTTCGACACTGATGATATCATATTCACGGCAGTACTACGTTCTC 121

Qy 41 ProLeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60
Db 122 CCCCTCATCTGGGCGCCTACAGTGGCGGCTAACTCTCGTCTCCGCTCGTGGCAACCG 181

Qy 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80
Db 182 TGTCCCTCTTTATPGACAGAGCGTTTCAGAGGTCAACAGGGCGCATTCCTCGTGAATTC 241

Qy 81 SerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMet 100
Db 242 TCAAACTGGAGGTCAGAGTTGGTTCGTCGCCGAGAGAGAACCTCAACATCAAGATG 301

Qy 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSer 120

```

```

Db 302 GATGTCAACCTACGATCTCGCTCAGCTAGCTATTATGGTGGCTCACTCCAGCCCCAGT 361
Qy 121 ProTrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSer 140
Db 362 CCCTGGAGGTCGTTGTTTCATAGCGGCTGGCTTAAGCCAGCAAGCTGGAGGAGATTCG 421
Qy 141 SerArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheVal 160
Db 422 TCGAGGAGTTTCTTCAGATCAAGAAACTGAAGCCAACTTAACGCTTACAGTTTGTGA 481
Qy 161 PheCysSerGluGlyAsnAspCysIleAspValGlyLysAsnGluGluGlyGlyValArg 180
Db 482 TTCTGTAGTAGGCTAAGGATTCATGATGTCGTTAAACACGAGGAAGTGGCGTTCGG 541
Qy 181 GlyLeuValLeuGlySerThrProPheAlaThrProPheGluValValPheValLys 200
Db 542 GGTGGTGGTTTGGTCTCTACGCCACCATTCGTCATCCCAAGACTATGCTATTATC 601
Qy 201 AlaThrGlyThrAspThrSerSerLysThrMetSerIleIle 214
Db 602 GCTACTGGACAGACACTTCATCCAGACTATGCTATTATC 643

RESULT 2
AY081323          845 bp      mRNA      linear      PLN 24-MAR-2002
LOCUS             Arabidopsis thaliana putative trypsin inhibitor (Atlg73260) mRNA,
DEFINITION        complete cds.
ACCESSION         AY081323
VERSION           AY081323.1 GI:19698952
KEYWORDS          FLI_CDNA.
SOURCE            Arabidopsis thaliana.
ORGANISM          Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE         1 (bases 1 to 845)
AUTHORS           Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
                  Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
                  Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
                  Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
                  Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
                  Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
                  Direct Submission
                  Submitted (26-FEB-2002) DNA Sequencing and Technology Center,
                  Stanford University, 855 California Avenue, Palo Alto, CA 94304,
                  USA
                  e-mail for correspondence: arabesequence.stanford.edu

TITLE             RIKEN Genomic Sciences Center (GSC) members carried out the
JOURNAL           collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
                  Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
                  Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
                  Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W.
(SSP/Stanford) contributed equally to this work as PIs.

FEATURES          Location/Qualifiers
                   1..845
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="RAFL06-11-B11 (R17231)"
                     /note="This clone is in pBluescript
                     ecotype: Columbia"

```

```
gene 1..845
/ gene="At1g73260"
/ note="Tl18K17.7"
CDS 134..781
/ gene="At1g73260"
/ codon_start=1
/ product="putative trypsin inhibitor"
/ protein_id="AAL91212.1"
/ db_xref="GI:19698953"
/ translation="MTTKTMNPKFYLVALTAVLASNAYGAVVDIGNAMPHESVY
VLPIRGSGGLTAGRGQCPDIQVSESEDEGIPVRFNMRLKVAEVPESQNLN
IETDVGATIGSNIRVGEDEHKKQYFVAGPKPEFGQDSLKSFKEKSGEDAY
KFVFCPTCDGSGNPKCSVDYGFIDELGVRRLALSDKPLVMEKKNATVSSKTM"
BASE COUNT 228 a 195 c 186 g 236 t
ORIGIN

Alignment Scores:
Pred. No.: 4,1e-39 Length: 845
Score: 506.50 Matches: 117
Percent Similarity: 63.59% Conservative: 21
Best Local Similarity: 53.92% Mismatches: 64
Query Match: 44.47% Indels: 15
DB: 8 Gaps: 7

US-09-822-080b-2 (1-214) x AY081323 (1-845)
QY 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsn 20
Db 155 ATGAATCCTAAGTTTACTTGGTCTTCCCTTAACCGCGGTCTCGGCTCAACGCA--- 211
QY 21 AlaGlyProValLeuAspThrAspGlyAspIlePheAspGlySerTyrTyrValLeu 40
Db 212 TATGTGCGGTTGTAGACATGATGGAACGCCATGTTCACGAAAGTTACTACGTTCTC 271
QY 41 ProLeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60
Db 272 CCTGTCATCCGTGGC---CGAGGCGGAGCGCTGACTTAGCAGCGCGGGTGGCAGCCA 328
QY 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80
Db 329 TGTCTTACGATATCGTCAGGAATCTTCAGAGTTGATGAGGCGCATTCGCTAAATTC 388
QY 81 SerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMet 100
Db 389 TCAAACTGGAGGCTTAAGGTTGCGTTCGTCCTCCGAATCACAGAACTCAACATCGAAACA 448
QY 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSer 120
Db 449 GACGTCGAGCCACGATCTGATCAGTCACCTACTGCGCGGTGCGGTGAGTTGACCCAC 508
QY 121 ProTrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSer 140
Db 509 GAGAGGAAGCAGTACTTCTGTTGCTGCTCCAAAGCCAGAGGGTTCGGACAGATTCG 568
QY 141 SerArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheVal 160
Db 569 TTGAGAGTTCTTCAGATCGAGAAATCTGGAGAG-----GATGCTACAAAGTTTGTG 622
QY 161 Phe-----CysSerGluGlyAsn---AspCysIleAspValGly---LysAsn 174
Db 623 TTCCTGCTCCGAGACTTCGACTCTGGCAATCCAAATGCGAGCATGTCGGGATATTCATA 682
QY 175 GluGluGlyGlyValArgGlyLeuValLeuGlySerThrProProPheAlaThrProPhe 194
Db 683 GATCAACTTGGCGTTCGCTGCTGCTTAAAGCAT-----AAGCGCTTC 727
QY 195 GluValValPheValLysAlaThrGlyThrAspThrSerSerLysThrMet 211
Db 728 TTGGTTATGTTCAAAAAGACTAATGTGACCGCAAGTTTCTCGCAGACATAG 778

RESULT 3
AC010556 70836 bp DNA linear PLN 15-JUN-2001
LOCUS AC010556
```

DEFINITION Arabidopsis thaliana chromosome 1 BAC Tl18K17 genomic sequence, complete sequence.

ACCESSION AC010556

VERSION AC010556.6 GI:12324309

KEYWORDS HTG.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana.

REFERENCE 1 (bases 1 to 70836)

AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J., Wu,D., Maiti,R., Bowmen,C.L., White,O., Nierman,W.C. and Fraser,C.M.

TITLE Arabidopsis thaliana chromosome 1 BAC Tl18K17 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 70836)

AUTHORS Lin,X. and Kaul,S.

TITLE Direct Submission

JOURNAL Submitted (15-SEP-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org

REFERENCE 3 (bases 1 to 70836)

AUTHORS Town,C.D. and Kaul,S.

TITLE Direct Submission

JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, ctown@tigr.org

COMMENT On Jan 19, 2001 this sequence version replaced gi:12280860. Address all correspondence to:at@tigr.org

BAC clone Tl18K17 is from Arabidopsis thaliana chromosome 1

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GenemarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/Genemark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/soflab/glimmerm.htm/glimmerm.html), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgl.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arfan Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES

Source

1..70836

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="1"

/clone="Tl18K17"

/complement(382..402)

/rpt\_family="AT-rich"

516..1979

repeat\_region

gene

note="predicted by genscan"

join(<516..836,1121..1263,1342..1526,1807..>1979)

/gene="Tl18K17.1"

join(516..836,1121..1263,1342..1526,1807..1979)

/gene="Tl18K17.1"

/codon\_start=1

/product="hypothetical protein; 516-1979"

/protein\_id="AAG52120.1"

/db\_xref="GI:12324311"

/translation="MHMEAAEIVHSAARFEAKMGKEDEAVEEIRRMSSGGYVIVVG"

mRNA

CDS







```

Pred. No.: 9,21e-14 Length: 845
Score: 239.50 Matches: 75
Percent Similarity: 50.46% Conservative: 34
Best Local Similarity: 34.72% Mismatches: 82
Query Match: 21.03% Indels: 25
DB: 8 Gaps: 11

US-09-822-080B-2 (1-214) x AF128268 (1-845)

QY 3 ProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThraAlaAsnAlaGly 22
Db 73 CCTCTCTTCTGCTTC-----CTGGCTCTTTCAGAGATGTTGAA 111
QY 23 ProValLeuAspThrAspGlyAspIleIlePhe---AspGlySerTyrTyrValLeuPro 41
Db 112 CAAGTTGTGGACATAAGTGGCAACCCATTTCCAGGTGGCACATATATACATATGCA 171
QY 42 LeuIleTyrGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsn---Gln 60
Db 172 TCAACTTGGGGCGCTCCCGGTGGTGAATACTA---GGCGGAGCTGGAACACTCAAAC 228
QY 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80
Db 229 TCCAGATTAAGTTTTCAGAGATTACTCAGAAATCTTCCGTGGCACACAGTCAAAATTC 288
QY 81 SerAsnTrpArgSerArgValGlyPheValProGluGluAsnLeuLysIleLysMet 100
Db 289 AGCATACCTGGGATAAGCCCTGGAATCATCTTACAGGTACTCCACITCAATCGAGTTC 348
QY 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrpTyrValThrProAlaProSer 120
Db 349 CGAGAGAAACCT---TATTGTGCTGAATCTCCAAATGGTGGCGTTTGTGGCAATGAA 405
QY 121 ProTyrArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSer 140
Db 406 ATCCAAAGGCGATGTGGGTATTGGTCTGTAAGGTCTATCTGTGGTCAACAAACATTT 465
QY 141 SerArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheVal 160
Db 466 AGTGCACACA---TTTAGCATTCAGAAATATAAATTT-----GGTACAACTTTGTG 513
QY 161 PheCysSerGluGlyAsn---AspCysIleAspValGly-----LysAsnGlu 175
Db 514 TTCTGTATCACTGGCTCAGGCACCTTGTATTAGATATTGGAAGGTTTGATGCGCAAAATGGT 573
QY 176 GluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGlu 195
Db 574 GAGGAGGAACGCT-----TTGAATCTCACTGAGCATGAGGCC-----TTCGAC 618
QY 196 ValValPheValLysAlaThrGlyThrAspThrSerSerLysThrMet 211
Db 619 ATTGTTTTCATAGAAGCTTCTAAGGTTTGATGGAATTATCAAGTCCGTA 666

RESULT 7
LEU07076 791 bp DNA linear PLN 07-OCT-1998
LOCUS Lycopersicon esculentum Lemir gene, complete cds.
DEFINITION U70076
ACCESSION U70076
VERSION U70076.1 GI:2654439
KEYWORDS
SOURCE Lycopersicon esculentum.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 791)
Brenner, E.D., Lambert, K.N., Kaloshian, I. and Williamson, V.M.
Characterization of Lemir, a root-knot nematode-induced gene in
tomato with an encoded product secreted from the root
Plant Physiol. 118 (1), 237-247 (1998)
98404251
PUBMED 9733543

REFERENCE
AUTHORS Lambert, K.N., Brenner, E.D., Kaloshian, I. and Williamson, V.M.
TITLE Direct Submission
JOURNAL Submitted (09-SEP-1996) Nematology, University of California at
Davis, Hutchison Hall, Davis, CA 95616, USA
FEATURES
Location/Qualifiers
source 1..791
/organism="Lycopersicon esculentum"
/strain="VFNT Cherry"
/db_xref="taxon:4081"
47..664
/notes="soybean Kunitz trypsin inhibitor family; miraculin
homolog; expressed in roots"
/codon_start=1
/product="Lemir"
/protein_id="AAC63057.1"
/db_xref="GI:2654440"
/translation="MKINLFFPELILATSPNLSLSSAAESPVEVDIDGKILRTGVD
VYLIVVRGGGLTMDISIGDKMCPDVAVOEHNEIDQGLPLTFTFPVDPKKGVIRES
DLNIFPANSICVQTTQWKLDDFDDETFQYFIFLGGDQNGPGETISNFKLEKYDRD
YKLLCYPTVCDKFCVKICRDIGIFIDQGVRLALSDVPEKVMFKKA"
BASE COUNT 249 a 151 c 150 g 241 t
ORIGIN
Alignment Scores:
Pred. No.: 1.47e-13 Length: 791
Score: 237.00 Matches: 65
Percent Similarity: 50.72% Conservative: 40
Best Local Similarity: 31.40% Mismatches: 84
Query Match: 20.81% Indels: 18
DB: 8 Gaps: 8

US-09-822-080B-2 (1-214) x LEU07076 (1-791)
QY 3 ProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThraAlaAsnAlaGly 22
Db 71 CCCTCTCTCATTTTAGCCATCTCTTTAACTCTTACTCTCTCAGCTGCTGAATCCCCA 130
QY 23 Pro---ValLeuAspThrAspGlyAspIleIlePheAspGly---SerTyrTyrValLeu 40
Db 131 CCGGAAGTAGTCACATTTGATGAAAGAGATACTCCGTACAGGCGTCGATTACTACATTTA 190
QY 41 ProLeuIleTyrGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60
Db 191 CCGGTGGTACCTGCG---AGAGGTGGTGGACTCACCATGACAGATATTGGTGAATAATG 247
QY 61 CysProLeu---PheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79
Db 248 TGTCCACTTGATGCTGTGTGCCAAGAACATAATGAGATTGATCAAGGCTCCCATTTGACA 307
QY 80 PheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnIleLys 99
Db 308 TTCACACCCGCTTGATCCAAAGAAAGGTGTGATTCGCGAATCGCACTGATTAAACATCAT 367
QY 100 MetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpTyrValThrProAlaPro 119
Db 368 TTTTTCAGCAAAAT---TCTATTGTGTTCACAAACACACATGGAAGCTAGATGACTTTGAT 424
QY 120 SerProTrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAsp 139
Db 425 GAAACACACAGGACATACTTTATTACGCTTGGTGGAGATCAAGGAACCCCTGGTGTGGAA 484
QY 140 SerSerArgSerPhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPhe 159
Db 485 ACTATAAGTAATTTGGTTCAGATTGAGAATACGAT-----CGCGATTATTAAGCTG 535
QY 160 ValPheCysSerGluGlyAsnAsp-----CysIleAspValGlyLysAsn 174
Db 536 CTATATTGCTTACATATGATGTGATTTCTCAAGATCATTTGTAGAGATTTGGCATATTC 595
QY 175 GluGluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPhe 194
Db 596 ATTCAAGGTGAGCTTAGACGTTTGGCTTTGAGTGAT-----GTCCCATTT 640

```





PASPHGGGLTMAFVPLPCLLVAQETDERKGFPPVRFETPWGGAAPEDRTIRVSTDV  
RIRNNAATICVQSTEMHWGDEPLTGARRVTGLIPSPSGRANAFRVERKGGYKLV  
SCRSQDLGVSRDAGRAWLGASQPPHVVFKRPSPE"

BASE COUNT 175 a 303 c 332 g 157 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,07e-12 Length: 967  
Score: 226.00 Matches: 71  
Percent Similarity: 48.77% Conservative: 28  
Best Local Similarity: 34.98% Mismatches: 74  
Query Match: 19.84% Indels: 30  
DB: 8 Gaps: 13

US-09-822-080B-2 (1-214) x AB075524 (1-967)

QY 8 LeuLeuAlaPheThrThrValLeuAlaAlaThraAlaAlaGlyProValLeuAspThr 27  
Db 263 CTCTGGCCATCTCC-----TTCATGTCAGCGCGCGCCGCCCGCTGTACGACACG 316  
QY 28 AspGly---AspIleIlePheAspGlySerTyrTyrValLeuProLeuIleTrpGlyPro 46  
Db 317 GAGGGCCAGCGCTGAGCGCCGACGGAGCTACTACGTCTCTCCCGGTAGCCCGGC--- 373  
QY 47 ThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysProLeuPheIleGly 66  
Db 374 CACGGAGGGGCTCAGCATG---GCGCCCGCGCTGCTCCCGCCGCTCTCTGTGGCG 430  
QY 67 GlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSerArg 86  
Db 431 CAGGACGCGAGCGCGCGCAAGGGGTTCCTCCGCTTACCCCTTACCCCGTGGCGGCC 490  
QY 87 ValGlyPheValProGluGluAsnLeuAsnIleLysMetAspValCldu----- 103  
Db 491 GCG-----GCGCGGAGGAGGACCATCCCGCTCTCGACGAGCTCGGATCGCTTC 544  
QY 104 -----ProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSerPro 121  
Db 545 AACCGCGCGCATGCTGCTGAGTCCACCGAGTGGCATGTGCGGCGACGCGCTCAGC 604  
QY 122 TrpArgSerLeuPheIleAlaAla-----GlyProLysProGluAlaGlyGlyGlu 138  
Db 605 GGGGCGCGCGGTGTGACGGGGCGGTGATCGGCGCGAGCCCG---AGCGGGCGGAG 661  
QY 139 AspSerArgSerPhePheGlnIleLysLysThrGluAlaLysAsnAlaTyrLys 158  
Db 662 AACGCG-----TTCGCGCTGGAGAGTACGCG-----CGTGGGTACAG 700  
QY 159 PheValPheCysSerGluClyAsnAspCysIleAspValGlyLysAsnGluGlyGly 178  
Db 701 CTGGTGTCTGACGGGAC-----TCGTGCCAGGACCTGGGCGTCAAGGAGCGCGCG 754  
QY 179 ValArgGlyLeuValGlySerThrProPheAlaThrProPheGluValPhe 198  
Db 755 -----CGGGCGTGGTGGGCGGAGCGCG-----CCFACCTGTGTGTCTTC 799  
QY 199 ValLysAla 201  
Db 800 AAGAAGGCC 808

RESULT 10  
OSJN00221/c  
LOCUS  
DEFINITION  
Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBA0038010,  
complete sequence.  
ACCESSION AL663019  
VERSION AL663019.2 GI:21912531  
KEYWORDS HTG.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

# REFERENCE AUTHORS

1 Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q. Q., Hu, X.,  
Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J., Zhang, L.,  
Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y., Li, C.,  
Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K., Zhou, B.,  
Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P., Fu, G.,  
Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F., Tu, Y. F.,  
Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X. Y.,  
Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W., Wang, L. J.,  
Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L., Zhu, F. H. and  
Hong, G. F.

# TITLE JOURNAL

Direct Submission  
Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,  
Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,  
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:  
bhan@ncgr.ac.cn

# REMARK

Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
clone: OSJNBA0038010.

# COMMENT

On Jul 19, 2002 this sequence version replaced gi:17998660.  
Web site: <http://www.ncgr.ac.cn>  
----- Summary Statistics  
Assembly program: phrap  
-----

# NOTE:

This is a complete sequence.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Egenes (<http://www.softberry.com/>),  
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM  
(<http://genemark.biology.gatech.edu/Genemark/>), TRNAScan-SE (Sean  
Eddy, <http://genome.wustli.edu/eddy/TrnScan-SE/>), searches of the  
complete sequence against NCBI none redundant protein database (nr)  
(<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.

# FEATURES

Location/Qualifiers  
source  
1..141545  
/organism="Oryza sativa"  
/variety="Nipponbare"  
/sub-species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="4"  
/clone="OSJNBA0038010"  
/clone\_lib="CUGI-OSJNBA"  
BASE COUNT 37180 a 32911 c 32798 g 38656 t  
ORIGIN

# Alignment Scores:

Pred. No.: 7,22e-10 Length: 141545  
Score: 226.00 Matches: 71  
Percent Similarity: 48.77% Conservative: 28  
Best Local Similarity: 34.98% Mismatches: 74  
Query Match: 19.84% Indels: 30  
DB: 8 Gaps: 13

US-09-822-080B-2 (1-214) x OSJN00221 (1-141545)

QY 8 LeuLeuAlaPheThrThrValLeuAlaAlaThraAlaAlaGlyProValLeuAspThr 27  
Db 74731 CTCTGGCCATCTCC-----TTCATGTCAGCGCGCGCCCGCTGTACGACACG 74678  
QY 28 AspGly---AspIleIlePheAspGlySerTyrTyrValLeuProLeuIleTrpGlyPro 46  
Db 74677 GAGGGCCAGCGCTGAGCGCGGAGCTACTAGCTCTCCCGGTAGCCCCCGCG--- 74621  
QY 47 ThrGlyGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysProLeuPheIleGly 66  
Db 74620 CACGGAGGGGCTCAGCATG---GCGCCCGCGGTGCTCCCTGCGCTCTCTGTGGCG 74564  
QY 67 GlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSerArg 86  
Db 74563 CAGGAGCGGACGAGCGCGCAAGGGGTTCCTCGTGGCTTACCCGCTGGCGCGCGC 74504  
QY 87 ValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAspValGlu----- 103  
Db 74503 GCG-----GCGCGGAGGAGACGACCATCGGTCTCTCGACCGAGCTCCGCTTC 74450



```

BASE COUNT      247 a      183 c      218 g      268 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      4.17e-12      Length:      917
Score:          222.50      Matches:      72
Percent Similarity: 48.28%      Conservative: 26
Best Local Similarity: 35.47%      Mismatches: 66
Query Match:    19.53%      Indels:      40
DB:              6      Gaps:      10

US-09-822-080B-2 (1-214) x I65559 (1-917)
QY 3 PrometPheTyrPheLeuLeuAlaPheThr-----ThrValLeuAlaAlaThrAlaAsn 20
Db 37 CCGTAGTTTACTC-CTCTTCGCCCTTCACATCAAAATCATATTTCTTTGGGGTACGCAAC 95
QY 21 Ala-----GlyProValLeuAspThrAspGlyAspIleIlePheAspGly---SerTyr 37
Db 96 GCTGCAAACTCTCCTGCTTGACACTGATGGTGATGAGCTCCAACTGGGGTTCATAT 155
QY 38 TyrValLeuProLeuIleTyrPglyProThrGlyGlyLeuThrLeuValSerArgArg 57
Db 156 TACGCTTTCATCGATATCGGGTCTGGGGTGGAGGCTAGCCCTAGGAAGGGCTACA 215
QY 58 GlyAsnGlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIlePro 77
Db 216 GGTCAAAGCTGCCAGAAATGTTGTCAAAGACGATCCGACCTTGACAATGGTACTCCT 275
QY 78 ValLysPheSerAsnTyrArgSerArgValGlyPheValProGluGluGluAsnLeuAsn 97
Db 276 GTAACTTTTCAAAATCGGATAGCAAGATGATGTTGTCGGGTACTACTGATGATAAC 335
QY 98 IleLys---MetAspValGluProThrIleCysAlaGlnSerAlaTyr-----112
Db 336 ATAGAGTTGCTTCCCATCAGACAGACTCTGCTCAACGTCACACTGTCTGGAGGCTTGC 395
QY 113 -----TyrPrpValThrProAlaProSerFroTyrPrpArgSer 124
Db 396 AATTATGCAACTCGCGAGGCAAAATGGTGGGGTGACA-----431
QY 125 LeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArgSerPhe 144
Db 432 -----ACTGATGGGGTTAAAGGTGAACCT---GGTCTTAACACTTTGTGCAAGTTGG 479
QY 145 PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValIleCysSerGlu 164
Db 480 TTTAAGATTGAGAAG-----GCCGGAGTACTTCGGTTACAAATTCAGGTTCCTCTTCT 533
QY 165 GlyAsnAsp-----CysIleAspValGlyLys-----AsnGluGluGly 177
Db 534 GTCGTGATTCGTGGCAACACTTTATGACCGATATTGGAAGACATTCACATGATGATGGA 593
QY 178 GlyValArg 180
Db 594 CAAATACGT 602

RESULT 13
TC21KD
LOCUS
DEFINITION
T.cacao mRNA for 21 kDa seed protein, homolog. to soybean trypsin inhibitor.
ACCESSION X56509
VERSION X56509.1 GI:21908
KEYWORDS protease inhibitor; trypsin inhibitor.
SOURCE Theobroma cacao.
ORGANISM Theobroma cacao
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Theobroma.
REFERENCE 1 (bases 1 to 930)
AUTHORS McHenry, L.

```

---

```

TITLE Direct Submission
JOURNAL Submitted (08-NOV-1990) L. McHenry, THE PENNSYLVANIA STATE UNIV,
215 BORLAND LABORATORY, UNIVERSITY PARK, PA 16802, USA
REFERENCE 2 (bases 1 to 930)
AUTHORS Tai, H., McHenry, L., Fritz, P.J. and Furtek, D.B.
TITLE Nucleic acid sequence of a 21 kDa cocoa seed protein with homology to the soybean trypsin inhibitor (Kunitz) family of protease inhibitors
JOURNAL Plant Mol. Biol. 16 (5), 913-915 (1991)
MEDLINE 91316229
PUBMED 1859871
REFERENCE 3 (bases 1 to 930)
AUTHORS Tai, H., McHenry, L., Fritz, P.J. and Furtek, D.B.
TITLE Nucleic acid sequence of a 21kD cocoa seed protein with homology to the soybean trypsin inhibitor (Kunitz) family of protease inhibitors
JOURNAL Unpublished
COMMENT AcNo is misprinted in publication as x54509.
FEATURES
    Location/Qualifiers
        1..930
            /organism="Theobroma cacao"
            /db_xref="taxon:3641"
            /tissue_type="seed"
            /clone_lib="lambda gt10"
            1..930
            /gene="asp"
            1..930
            /gene="asp"
            /evidence=experimental
            31..696
            /gene="asp"
            /note="homol. to soybean trypsin inhibitor"
            /codon_start=1
            /product="21 kDa seed protein"
            /protein_id="CAA39860.1"
            /db_xref="GI:21909"
            /db_xref="SWISS-PROT:P32765"
            /translation="MKTATAVVLLLFATFSKVFVGVAANSPVLDTDGDDELQGVQ
            YVLSISGAGGGLALGRATGSCPEIVVORSDLDNGTPVIFSNADSKDDVVRVST
            DVNIERPVRDLRCSTSTVWRLDNDNSAGKWWTTDGVKGEPGNTLCSWFKIEKAG
            VLGYKRFPCPSVCDSCCTTLCSDIGRHSDDGQIRLALSDNEWAMFFKASKRTIKQVVN
            AKH"
            sig_peptide
            31..108
            /gene="asp"
            109..693
            /gene="asp"
            /product="21 kDa seed protein"
            /note="homol. to soybean trypsin inhibitor"
            930
            /gene="asp"
BASE COUNT      250 a      191 c      216 g      267 t      6 others
ORIGIN

Alignment Scores:
Pred. No.:      4.24e-12      Length:      930
Score:          222.50      Matches:      72
Percent Similarity: 48.28%      Conservative: 26
Best Local Similarity: 35.47%      Mismatches: 66
Query Match:    19.53%      Indels:      40
DB:              8      Gaps:      10

US-09-822-080B-2 (1-214) x TC21KD (1-930)
QY 3 PrometPheTyrPheLeuLeuAlaPheThr-----ThrValLeuAlaAlaThrAlaAsn 20
Db 47 CCGTAGTTTACTC-CTCTTCGCCCTTCACATCAAAATCATATTTCTTTGGGGTACGCAAC 105
QY 21 Ala-----GlyProValLeuAspThrAspGlyAspIleIlePheAspGly---SerTyr 37
Db 106 GCTGCAAACTCTCCTGCTTGACACTGATGGTGATGAGCTCCAACTGGGGTTCATAT 165
QY 38 TyrValLeuProLeuIleTyrPglyProThrGlyGlyLeuThrLeuValSerArgArg 57

```

```

Db 166 TACGCTCTTGTGTCATCGATATCGGGTGTGGGGTGGAGGGCTAGCCCTAGGAAGGGCTACA 225
Qy 58 GlyAsnGlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIlePro 77
Db 226 GGTCAAGAGTCCCAAGAAATGTTGTCCAAGAGGATCCGACCTTGACATGTTACTCCT 285
Qy 78 VallysPheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsn 97
Db 286 GTAATCTTTTCAATGGGATACGAAGATGATGTTGTCCGCGTATCTACTGATGTAAC 345
Qy 98 IleLys---MetAspValGluProThrIleCysAlaGlnSerAlaIlyr----- 112
Db 346 ATAGATGCTGTTCCATCAGACAGACACTCTGCTCAACGCTCAACTGTGTGGAGCTTGAC 405
Qy 113 -----TTPTrpValThrProAlaProSerProTrpArgSer 124
Db 406 AATTATGACAACCTCGGAGCAATGGTGGTGACA----- 441
Qy 125 LeuPheIleAlaGlyProLysProGluAlaGlyGlyGluAspSerArgSerPhe 144
Db 442 -----ACTGATGGGTTAAAGGTGAACCT---GGTCTTAACACTTTGTGCAGTTGG 489
Qy 145 PheGlnIleLysLysThrGluAlaLysLeuAsnAlaIlyrLysPheValPheCysSerGlu 164
Db 490 TTTAAGATTGAGAAG-----GCCGGAGTACTCGGTTACAAATTCAGGTTCTGCTCTTCT 543
Qy 165 GlyAsnAsp-----CysIleAspValGlyLys-----AsnGluGluGly 177
Db 544 GTCTGTGATTCGTCACACTTTATCGCGGATATTTGGAAGACATTCAGATGATGGA 603
Qy 178 GlyValarg 180
Db 604 CAATACGT 612

RESULT 14
LOCUS CAR276263
DEFINITION Cicer arietinum mRNA for alpha-fucosidase, clone CanFuc-2.
ACCESSION AJ276263
VERSION AJ276263.3 GI:14161087
KEYWORDS alpha-fucosidase.
SOURCE chickpea.
ORGANISM Cicer arietinum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae;
Cicer.
1 (bases 1 to 792)
Duplico.B., Jimenez,T. and Labrador,E.
A second alpha-fucosidase is expressed in chickpea epicotyls
Unpublished
2 (bases 1 to 792)
Labrador,E.
Direct Submission
Submitted (29-FEB-2000) Labrador E., Dpto. Fisiologia Vegetal,
Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
Reina s/n, E-37007, SPAIN
3 (bases 1 to 792)
Labrador,E.
Direct Submission
Submitted (16-MAY-2001) Labrador E., Dpto. Fisiologia Vegetal,
Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
Reina s/n, E-37007, SPAIN
4 (bases 1 to 792)
Labrador,E.
Direct Submission
Submitted (18-MAY-2001) Labrador E., Dpto. Fisiologia Vegetal,
Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la
Reina s/n, E-37007, SPAIN
On May 20, 2001 this sequence version replaced gi:14148962.
Location/Qualifiers

```

```

source
1. 792
/organism="Cicer arietinum"
/cultivar="Castellana"
/db_xref="taxon:3827"
/clone="CanFuc-2"
/tissue_type="etiolated epicotyls"
/clone_lib="CAN-5"
/dev_stage="seedling"
/country="Spain"
/note="age 5 days"
6. 653
/EC_number="3.2.1.51"
/codon_start=1
/product="alpha-fucosidase"
/protein_id="CAB76907.3"
/db_xref="GI:14161088"
/db_xref="sptrembl:Q9M327"
/translation="MKQSFLLSLFVLLNLSLAFSNEDVEQVLDINGNPFPGGKY
YILPAIRGPPGGVRLDKTGDSECPVTVLQDYKEVINGLPVKFVIGLSPIIITGTP
IEIEFTKPNCAESSKWLIEDDTIDKACIGIGPENYSKQILSGTFNLIQYGSFG
YKLGFCVKGSPICLDIGRIDNDEGRRLNLTETHEAFRVFVDASSYEDGIVKSVV"
BASE COUNT 241 a 116 c 152 g 283 t
ORIGIN
Alignment Scores:
Pred. No.: 5.44e-12 Length: 792
Score: 220.50 Matches: 68
Percent Similarity: 51.72% Conservative: 37
Best Local Similarity: 33.50% Mismatches: 85
Query Match: 19.36% Indels: 13
DB: 8 Gaps: 9
US-09-822-080B-2 (1-214) x CAR276263 (1-792)
Qy 4 MetPheTyrPheLeuLeuAlaPheThrThrValLeuAlaAlaThrAlaAsnAlaGlyPro 23
Db 36 CTCCTTTGTTTCTTACTCAATCTTTTCACTAGCTTTC---TCAAACGAAGATGTTGAACAA 92
Qy 24 ValLeuAspThrAspGlyAspIleIlePhe---AspGlySerTyrTyrValLeuProLeu 42
Db 93 GTACTAGACATAAATGGTAACTCTATTTTCCAGGTGGCAAAATACTATATTTTCCCGCGCA 152
Qy 43 IleTrpGlyProThrGlyGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysPro 62
Db 153 ATTCGTGGACCTCCAGGTGGAGGAGTCAGATTAGACAAAACACTGCTGATTGAGAGTGCCA 212
Qy 63 LeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82
Db 213 GTTACCCTCTTACAAGACTACAAAGAGTTATCAATGGTTTACCAGTAAATTTGTTATA 272
Qy 83 TrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspVal 102
Db 273 CCAGATAAGTCTCTGGTAAATAATTTACTGGTACACCAATTTGAAATCGAGTTCACATAA 332
Qy 103 GluProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSerProTrp 122
Db 333 AAGCTAAT---TGTGCTGAATCATCAAAATGGTTGATATTTGTCGACGATACACTATCGAC 389
Qy 123 ArgSerLeuPheIleAlaGlyProLysProGluGluGlyGlyGluAspSerArg 142
Db 390 AAAGCTTGATTTGGTATTTGGTGTCTCTGAAATTTATTCGGGTAAACAAACATTTGAGTGGC 449
Qy 143 SerPhePheGlnIleLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCys 162
Db 450 ACA---TTTAAATATTCAAAATAATGATGATCTGGATTT---GGTTATAAGCTTGGATTTGT 503
Qy 163 SerGlyGlyAsnAsp---CysIleAspValGlyLys-----AsnGluGluGlyVal 179
Db 504 GTTAAGGGTTCACCAATTTGTTGGATATTTGGAAGATATCATATATGATGAAGGTGGAAGA 563
Qy 180 ArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGluValValPheVal 199
Db 564 CGT-----TTGAATTTTGACTGAACATGAAGCT-----TTTGAAGTTGTTTGT 608

```

QY 200 LysalaThr 202  
 Db 609 GATGCTTCT 617  
 RESULT 15  
 AF356620  
 LOCUS  
 DEFINITION Theobroma bicolor clone 5-2 trypsin inhibitor gene, partial cds.  
 ACCESSION AF356620  
 VERSION AF356620.1 GI:19171720  
 KEYWORDS  
 SOURCE Theobroma bicolor.  
 ORGANISM Theobroma bicolor  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Theobroma.  
 REFERENCE 1 (bases 1 to 465)  
 AUTHORS Silva, C.R.S. and Figueira, A.  
 TITLE Phylogeny of genus Theobroma and Herrania (Sterculiaceae) based on sequences of the nuclear gene trypsin inhibitor  
 JOURNAL unpublished  
 REFERENCE 2 (bases 1 to 465)  
 AUTHORS Silva, C.R.S. and Figueira, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-2001) LAMP - A/C Antonio Figueira, Centro de Energia Nuclear na Agricultura, Av. Centenario, 303 - CP: 96, Piracicaba, SP 13.400-970, Brazil  
 FEATURES  
 source  
 1. .465  
 /organism="Theobroma bicolor"  
 /db\_xref="taxon:108878"  
 /clone="5-2"  
 /tissue\_type="leaf"  
 /dev\_stage="mature tree"  
 /country="Brazil: Marituba, PA; 1 deg 12'S; 49 deg 30'W"  
 /note="sample from the Basil Bartley Theobroma collection of the 'Comissao Executiva do Plano da Lavoura Cacaueira - CEPLAC'."  
 mRNA  
 <1..>465  
 /product="trypsin inhibitor"  
 CDS  
 <1..>465  
 /note="seed protein; protease inhibitor; similar to cocoa trypsin inhibitor"  
 /codon\_start=3  
 /product="trypsin inhibitor"  
 /protein\_id="AA1856S5.1"  
 /db\_xref="GI:19171721"  
 /translation="VLDTDGDELQTVQYVYVSSIWAGGGGLALGRATGQSCPEIYV  
 QRQDFDRGTVPVIFSNADSEGVVSTDIIEFVPICDRLCTTSTWKLNDYNSAG  
 QWVTTDGDGKGPNNLTWFKIKESGVMSYKRFPGVSPFDCSCTTLCSDIG"  
 BASE COUNT 118 a 89 c 129 g 129 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 5,6e-12 Length: 465  
 Score: 217.50 Matches: 60  
 Percent Similarity: 47.62% Conservative: 20  
 Best Local Similarity: 35.71% Mismatches: 55  
 Query Match: 19.10% Indels: 33  
 DB: 8 Gaps: 7

US-09-822-080B-2 (1-214) x AF356620 (1-465)

QY 24 valLeuAspThrAspGlyAspIlellePheAspGly---SerTyrTyrValLeuProLeu 42  
 Db 3 GTGCTTGACACTGATGGTATGAGCTCCAAACTGGGGTTCAATATTACGTGTCATCA 62  
 QY 43 IleTrpGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCysPro 62  
 Db 63 ATATGGGGTCTGGGGTGGGGCTAGCTCTTGGAGGGGCTACAGGCTAAAGCTGCCCA 122  
 QY 63 LeuPheIleGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82

Db 123 GAAATGTGTTTCAAGACAATCCGACTTTGACCGTGGTACTCCTGTGATCTTTTCAAAT 182  
 QY 83 TrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLys---MetAsp 101  
 Db 183 GCGGATAGCGAAGATGGTGTGTCGCCGTATCTACCGATATAAACATAGAGTTCGTCC 242  
 QY 102 ValGluProThrIleCysAlaGlnSer----- 110  
 Db 243 ATCGGGGACAGACTCTGACGGAGCTCAACTGTGTGAAGCTTGACAAATTATGACAACTCG 302  
 QY 111 ---AlaTyrTrpTrpValThrProAlaProSerProTrpArgSerLeuPheIleAla 129  
 Db 303 GCAGGCCAATGGTGGGTGACA-----ACGGAT 329  
 QY 130 GlyProLysProGluAlaGlyGlyGluAspSerSerArgSerPheGlnIleLysLys 149  
 Db 330 GGGGATAAAGGTGAACCT---GGTCCTAACAAATTGACCAGTTGGTTAAGATTAAAG--- 383  
 QY 150 ThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSerGluGlyAsnAsp----- 167  
 Db 384 ---GAGTCCCGAGTAATGAGTTACAAATTTAGGTTCTGCTCTTCTGTGATTCGTGC 440  
 QY 168 -----CysIleAspValGly 172  
 Db 441 ACAACTTTATGCAGCGATATTGGA 464

Search completed: January 25, 2003, 20:24:31  
 Job time : 2950 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 25, 2003, 18:17:51 : Search time 259 Seconds  
(without alignments)  
1860.724 Million cell updates/sec

Title: US-09-822-080B-2  
Perfect score: 1139  
Sequence: 1 MNPFFYLLAFTTVAATAN.....EVVFKATGDTSSKTMSTII 214

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cqn2.1/USPTO.spool/US09822080/runat\_22012003.142637.4407/app\_query.fasta\_1.391  
-DB=N\_Geneseq.101002 -QPMF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THRESHOLD=1 -THRESHOLD=100 -THRESHOLD=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09822080.ecgn\_1.1.0 @runat\_22012003.142637.4407 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match %	Score	Length	DB ID	Description
1	506.5	44.5	792	24	ABN99075
2	214.5	18.8	916	13	AAQ20375
3	200.5	17.6	591	20	AAZ24282
4	194	17.0	539	16	AAQ03513
5	194	17.0	541	17	AAZ44489
6	192.5	16.9	519	20	AAZ24294
7	192.5	16.9	519	20	AAZ24295
8	192.5	16.9	519	20	AAZ24296
9	178	15.6	1092	21	AAZ33644
10	167	14.7	624	20	AAZ24283
11	161.5	14.2	2352	12	AAQ11490
12	160.5	14.1	954	7	AAQ0845
13	145	12.7	899	20	AAZ07003
14	145	12.7	899	22	AAZ07003
15	142	12.5	894	7	AAZ0844
16	137.5	12.1	860	21	AAQ38893
17	134	11.8	588	14	AAQ38954
18	126.5	11.1	630	14	AAQ38954
19	116.5	10.2	2490	12	AAQ11491
20	98.5	8.6	956	22	AAZ36455
21	96.5	8.5	605	24	ABO66345
22	96.5	8.5	990	21	AAZ55784
23	96.5	8.5	12249	21	AAZ55840
24	96.5	8.5	18331	21	AAZ55857
25	94.5	8.3	1403	20	AAZ86970
26	94.5	8.3	2456	22	AAZ06951
27	94.5	8.3	6641	22	AAZ06951
28	94	8.3	1825	23	ABL09271
29	94	8.3	3892	23	ABL09270
30	93	8.2	15446	24	AAZ25519
31	91	8.0	1362	24	ABK53020
32	91	8.0	1612	21	AAZ64692
33	91	8.0	1612	21	AAZ63946
34	91	8.0	1612	24	AAZ44747
35	91	8.0	4373	24	ABK53021
36	89.5	7.9	20311	23	ABL13796
37	89	7.8	987	24	ABQ90261
38	88	7.7	543	24	ABN61432
39	88	7.7	1004	21	AAZ35665
40	88	7.7	2783	22	AAZ64212
41	87.5	7.7	1443	20	AAZ25629
42	87.5	7.7	1443	20	AAZ25630
43	87.5	7.7	1801	22	AAZ01501
44	87.5	7.7	1801	23	ABL96945
45	87	7.6	647	22	AAZ23005

## ALIGNMENTS

RESULT 1

ABN99075

ID ABN99075 standard; DNA; 792 BP.

XX ABN99075;

AC ABN99075;

XX 01-AUG-2002 (first entry)

DT DT

XX Arabidopsis thaliana expressed polynucleotide SEQ ID NO 843.

DE Arabidopsis thaliana; plant; insecticide; fungicide; stress;  
KW disease; crop; thale cress; tolerance factor; insect; pathogen;  
KW nutrition; ds.

XX Arabidopsis thaliana.

XX US2002023281-A1.

PN







CC selected and sequences which cause mRNA stability are modified (see  
CC also AAX24294-96). The gene is prepared by PCR amplification (see  
CC also AAX24284-93). Also claimed is a plant containing the  
CC artificially synthesised gene, introduced via a vector comprising  
CC an expression cassette, a selectable marker, and the new gene  
CC linked to a plant promoter. Transgenic plants can be produced that  
CC are resistant to pests, especially lepidopteran insects.

XX  
SQ Sequence 591 BP; 165 A; 138 C; 129 G; 159 T; 0 other;

Alignment Scores:  
Pred. No.: 6,06e-13 Length: 591  
Score: 200.50 Matches: 57  
Percent Similarity: 50.25% Conservative: 44  
Best Local Similarity: 28.36% Mismatches: 75  
Query Match: 17.60% Indels: 25  
DB: 20 Gaps: 9

US-09-822-080B-2 (1-214) x AAX24282 (1-591)

Qy 7 PheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGlyProValLeuAsp 26  
Db TTTCTACTCTGTCATCATCTCACACCTACCATCTGCTGAGCCATTCGCTCGAT 87

Qy 27 ThrAspGlyAspIlePheAsp---GlySerTyrTyrValLeuProLeuLeuTyrGly 45  
Db TCTGAAGGTGAGTTGTCGAATGTTGGCAGCATACTACTTCTGTGGCAGATAGATGGCA 147

Qy 46 ProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCysProLeuPheIle 65  
Db TTTTGGGGGAGGAGTAAAGAGCAGCAGCAAGCAAGCAACATGCGCTCTTAACAGTG 204

Qy 66 GlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSer 85  
Db GTACGATCTCCCAATGAGGTCTGTAGGGGAAACCATTAAGGATCTCATCC---CAATTG 261

Qy 86 ArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspValGluProThr 105  
Db TTTTGGGGGAGGAGTAAAGAGCAGCAGCAAGCAAGCAACATGCGCTCTTAACAGTG 204

Qy 106 IleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrpArgSerLeu 125  
Db TTTTGGGGGAGGAGTAAAGAGCAGCAGCAAGCAAGCAACATGCGCTCTTAACAGTG 204

Qy 126 PheIleAlaAlaGlyProLysProGluAlaGly---GlyGluAspSerSerArgSerPhe 144  
Db GTTGAGGAGCAGCAGCAACACCTCTGTTAACTTACTGAGCTAAATCTACTAAATTC 408

Qy 145 -----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPhe 161  
Db GATTATCTATTCAAAATTCGAGAAAGTTACCTTAAGTTTCTCTCATACAAGCTTAAGTAC 468

Qy 162 CysSerGluGlyAsnAspCysIleAspValGly-----LysAsnGluGluGlyVal 179  
Db TGTGCCAAGAGGAGCAGCAGTAAAGGATATCGGATCTATAGGATCAGAAAGGATAGCCA 528

Qy 180 ArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGluValValPheVal 199  
Db TTTTGGGGGAGGAGTAAAGAGCAGCAGCAAGCAAGCAACATGCGCTCTTAACAGTG 204

Qy 200 Lys 200  
Db TTTTGGGGGAGGAGTAAAGAGCAGCAGCAAGCAAGCAACATGCGCTCTTAACAGTG 204

Qy 574 AAG 576

RESULT 4  
AAT03513  
ID AAT03513 standard; cDNA; 539 BP.  
XX  
AC AAT03513;  
XX  
DT 24-MAY-1996 (first entry)  
XX  
DE Erythrina caffra DE-3 Inhibitor cDNA, for expression in E.coli.

XX  
KW Recombinant; Erythrina caffra; DE-3 inhibitor; Immobilised;  
KW affinity; purification; serine protease; tissue plasminogen activator;  
KW chromatography; metal chelate; anion exchange; cation exchange;  
KW prokaryote; Escherichia coli; codon preference; codon bias; ds.  
XX  
OS Synthetic.

XX  
FH Key Location/Qualifiers  
FT misc\_feature 1..8  
FT /tag= a  
FT /function= multiple\_cloning\_site  
FT /note= "EcoRI restriction site"  
FT CDS 9..527  
FT /tag= b  
FT /product= DE-3 inhibitor  
FT /note= "initial Met residue only included in  
FT prokaryotic expression; coding sequence  
FT is optimised for expression in E.coli"  
FT misc\_feature 528..539  
FT /tag= c  
FT /function= multiple\_cloning\_site  
FT /note= "HindIII restriction site"

XX  
PN DE4424171-Al.  
XX  
XX 21-SEP-1995.  
XX  
XX 08-JUL-1994; 94DE-4424171.  
XX  
XX 16-MAR-1994; 94DE-4408939.  
XX  
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.  
XX  
XX Fischer S, Kohnert U, Stern A;  
XX  
XX WPI; 1995-359618/47.  
XX P-PSDB; AAR88277.  
XX  
XX Serine protease purific. on Erythrina caffra immobilised DE-3  
XX inhibitor - using recombinant material having higher specific  
XX affinity that natural inhibitor, partic. for purifying tissue  
XX plasminogen activator  
XX  
XX Claim 13; Page 8; 10pp; German.  
XX  
XX Recombinant DE-3 inhibitor obtained by expressing Erythrina caffra  
XX cDNA in prokaryotic or eukaryotic hosts and then purifying it by  
XX chromatography on anion or cation exchange resin or Ni chelate is  
XX found to have significantly greater specific affinity for serine  
XX proteases than the natural inhibitor. The recombinant inhibitor,  
XX immobilised on a suitable support, is used in an improved method  
XX for purifying serine proteases (esp. plasminogen activators) from  
XX mixtures of proteins. The present sequence encodes the DE-3  
XX inhibitor using codons preferred by E.coli.

XX  
SQ Sequence 539 BP; 144 A; 111 C; 168 G; 116 T; 0 other;

Alignment Scores:  
Pred. No.: 2.83e-12 Length: 539  
Score: 194.00 Matches: 58  
Percent Similarity: 50.52% Conservative: 39  
Best Local Similarity: 30.21% Mismatches: 67  
Query Match: 17.03% Indels: 28  
DB: 16 Gaps: 12

US-09-822-080B-2 (1-214) x AAT03513 (1-539)

Qy 24 ValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTyrTyrValLeuProLeu 42  
Db TTTATAGATGTAACGGCGAAGTGGTGCAGACGGCGGTACCTATTATCTGTCGCGCAG 74

Qy 43 IleTrpGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCysPro 62

Db 75 GTGTGGCG---CAGGCGCGCGCGTGCAGTGGCGGAAACCGCGAAGAACTGCCG 131  
QY 63 LeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82  
Db 132 CTGACCGTGTGCAGACCGCGAAGTGCAGGATGCGAAGCGATTCGTATTGAAAGC 191  
QY 83 TrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAspVal 102  
Db 192 ---CGTCTCGGTAGCGCGTTTATTCGGATGATGATAAAGTGCATATGGCTTTGCCGTAT 248  
QY 103 GluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTyr 122  
Db 249 GCGCCGAAA---TGCAGCGCGCGCGCGTGGTGGACCGTGGTGAAGATGAACAGGAAGC 305  
QY 123 ArgSerLeuPheIleAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArg 142  
Db 306 CTGAGCGTGAACCTGAGC-----GAGGATGAAGACACC 338  
QY 143 SerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPhe 159  
Db 339 CAGTTTGATTTACGCTTTAAATTTTGAACAGGTGAGCGTGCAGTGCATAGCTATAAAGC 398  
QY 160 ValPheCysSerGluGly-----AsnAspCysIleAspValGlyLysAsn---GluGlu 176  
Db 399 CTGTATTGC---GAAGCAACATGAATAATGCGGAGCATGGCATTAACCGTGATCAG 455  
QY 177 GlyGlyValArgGlyLeuValLeuLysSerThrProPheAlaThrProPheGluVal 196  
Db 456 AAGGCTATCTCTCTGTGGTGTG-----ACCGAAGATTAT-----CGCTGACCGGTG 503  
QY 197 ValPheValAlaThrGlyThrAspThrSerSer 208  
Db 504 GTGCTGAAAAA-----GATGAAGACGAGC 527

RESULT 5  
AAT44489  
ID AAT44489 standard; cDNA; 541 BP.  
XX AC AAT44489;  
XX AC AAT44489;  
XX DT 10-JUN-1997 (first entry)  
XX DE cDNA encoding protein with activity of DE-3 inhibitor.  
XX DE DE-3; serine protease; inhibitor; Erythrina caffra; purification; ds.  
XX KW Erythrina caffra.  
XX OS  
XX FH Location/Qualifiers  
XX FT 11..532  
XX FT /\*tag= a  
XX PN DE19512937-Al.  
XX XX  
XX PD 10-OCT-1996.  
XX XX  
XX PF 06-APR-1995; 95DE-1012937.  
XX XX  
XX PR 06-APR-1995; 95DE-1012937.  
XX XX  
XX PA (BOE ) BOEHRINGER MANNHEIM GMBH.  
XX PI Fischer S, Kohnert U, Stern A, Wozny M;  
XX WPI; 1996-456284/46.  
XX DR P-PSDB; AAW01434.  
XX XX  
XX PT Recombinant Erythrina caffra trypsin inhibitor - with modified  
XX N-terminus, for purifying serine protease enzymes  
XX PS  
XX Claim 7; Page 10-11; 14pp; German.

CC The present sequence encodes a polypeptide (AAW01434) with the activity  
CC of Erythrina caffra trypsin inhibitor DE-3. The protein is capable of  
CC reversibly and selectively binding serine proteases from protein mixts.  
CC and is useful for purifying such proteases (esp. recombinant tissue  
CC plasminogen activator) by a claimed process in which the protease is  
CC bound by the immobilised polypeptide, unbound components of the protein  
CC mixt. are removed, the immobilised polypeptide is sepd. from the soluble  
CC protease, and the protease is recovered.  
XX  
SQ Sequence 541 BP; 145 A; 113 C; 167 G; 116 T; 0 other;

Alignment Scores:  
Pred. No.: 2,84e-12 Length: 541  
Score: 194.00 Matches: 58  
Percent Similarity: 50.52% Conservative: 39  
Best Local Similarity: 30.21% Mismatches: 67  
Query Match: 17.03% Indels: 28  
DB: 17 Gaps: 12

US-09-822-080B-2 (1-214) x AAT44489 (1-541)  
QY 24 ValLeuAspThrAspGlyAspIlePheAsp---GlySerTyrTrpValLeuProLeu 42  
Db 17 TTATTAGATGTAACGCGAAGTGTGCAGACGCGGTACCTATTATCTGCTGCCGAG 76  
QY 43 IleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysPro 62  
Db 77 GTGTGGCG---CAGGCGCGCGGTGCAGTGGCGGAAACCGGATTCGTATTGAAAGC 193  
QY 63 LeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82  
Db 134 CTGACCGTGTGAGACCGCGAAGTGCAGCGATGATGATAAAGTGCATATGGCTTTGCCGTAT 250  
QY 83 TrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspVal 102  
Db 194 ---CGTCTGCGTAGCGCGTTTATTCGGATGATGATAAAGTGCATATGGCTTTGCCGTAT 250  
QY 103 GluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTyr 122  
Db 251 GCGCCGAAA---TGCAGCGCGCGCGTGGTGGACCGTGGTGAAGATGAACAGGAAGC 307  
QY 123 ArgSerLeuPheIleAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArg 142  
Db 308 CTGAGCGTGAACCTGAGC-----GAGATGAAAGACACC 340  
QY 143 SerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPhe 159  
Db 341 CAGTTTGATTTACGCTTTAAATTTTGAACAGGTGAGCGTGCATAGCTATAAAGC 400  
QY 160 ValPheCysSerGluGly-----AsnAspCysIleAspValGlyLysAsn---GluGlu 176  
Db 401 CTGTATTGC---GAAGCAACATGAATAATGCGGAGCATGGCATTAACCGTGATCAG 457  
QY 177 GlyGlyValArgGlyLeuValLeuLysSerThrProPheAlaThrProPheGluVal 196  
Db 458 AAGGCTATCTCTCTGTGGTGTG-----ACCGAAGATTAT-----CCGCTGACCGGTG 505  
QY 197 ValPheValLysAlaThrGlyThrAspThrSerSer 208  
Db 506 GTGCTGAAAAA-----GATGAAGACGAGC 529

RESULT 6  
AAX24294  
ID AAX24294 standard; DNA; 519 BP.  
XX AC AAX24294;  
XX AC AAX24294;  
XX DT 07-JUN-1999 (first entry)  
XX DE Synthetic gene encoding winged bean trypsin inhibitor WTI-1B.  
XX KW Trypsin inhibitor; winged bean; WTI-1B; chymotrypsin inhibitor;  
XX WCI-3; protease inhibitor; enzyme engineering; protein engineering;

TOZ VAGTUPTOMITTECSAIAA

CC This is the nucleotide sequence of an artificially synthesised gene

QY 102 VALUATION OF JOINTLY OWNED ASSETS

CC that codes for trypsin inhibitor WTI-1B (see AAW97839) of winged bean  
 CC (Psophocarpus tetragonolobus (L.) DC.). The synthetic gene is  
 CC designed for stable expression in rice plants. It was obtained by:  
 CC (i) comparing WTI-1B and winged bean chymotrypsin inhibitor WCI-3  
 CC (see AAW97840) amino acid sequences, and obtaining homologous amino  
 CC acid pairs (one each from WTI-1a and b); and (ii) selecting codons  
 CC from WCI-3 if the amino acid pair is the same, and from WTI-1B if  
 CC they are different. High usage codons are selected based on codon  
 CC usage of Fabaceae plants. Sequences which cause mRNA stability  
 CC can subsequently be modified (see AAX24296). The gene is prepared by  
 CC PCR amplification (see also AAX24284-93). A claimed synthetic gene  
 CC (see AAX24282) includes DNA encoding the WCI-3 signal peptide in  
 CC addition to the sequence encoding WTI-1B. Also claimed is a plant  
 CC containing the claimed synthetic gene. Transgenic plants can be  
 CC produced that are resistant to pests, especially lepidopteran  
 CC insects.

SQ Sequence 519 BP; 149 A; 110 C; 123 G; 137 T; 0 other;

#### Alignment Scores:

Pred. No.: 3.94e-12 Length: 519  
 Score: 192.50 Matches: 53  
 Percent Similarity: 50.81% Conservative: 41  
 Best Local Similarity: 28.65% Mismatches: 66  
 Query Match: 16.90% Indels: 25  
 DB: 20 Gaps: 9

US-09-822-080B-2 (1-214) x AAX24295 (1-519)

QY 23 ProValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTyrTyrValLeuPro 41  
 DB 4 CCATTCGTCGATTCGAAGTGGAGTTCGAAATGGTGGCAGATCTATCTCTGCCA 63  
 QY 42 LeuIleTyrGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCys 61  
 DB 64 GATAGATGGCA---CTCGGGGGAGGATAGACAGCAGCAGCAACGAAACATGC 120  
 QY 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSer 81  
 DB 121 CCTCTAACAGTGGTACGATCTCCCAATGAGTCTCTGTAGGGGAACCATTAAGGATCTCA 180  
 QY 82 AsnTyrArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAsp 101  
 DB 181 TCC---CAATGCGTTCAGGTTGATCCCGGATTACTCTCTAGTGCCTATTGGATTC--- 234  
 QY 102 ValGluProThrIleCysAlaGlnSerAlaTyrTyrTrpValThrProAlaProSerPro 121  
 DB 235 GCTAACCCCTCCAAAGTGT-----GACCTTCCTCCG 264  
 QY 122 TrpArgSerLeuPheIleAlaGlyProLysProGluAlaGly---GlyGluAspSer 140  
 DB 265 TGGTGGACTGTGTGGAGGACCAACCAACCCCTCTGTAAACTTAGTGAGCTAAAA 324  
 QY 141 SerArgSerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyr 157  
 DB 325 TCTACTAATTCGATATATATTAATTCGAGAAAGTTACTCTAAGTTTCCTCATAC 384  
 QY 158 LysPheValPheCysSerGluGlyAsnAspCysIleAspValGly-----LysAsnGlu 175  
 DB 385 AAGCTTAAGTACTGTGCCAAGGACACCTGTAAAGATATCGGGATTATAGGATCAG 444  
 QY 176 GluGlyGlyValArgGlyLeuValLeuGlySerThrProProPheAlaThrProPheGlu 195  
 DB 445 AAAGGATACGACGCTTGGTGGTACTGACGAA-----AACCCATTAGTG 489  
 QY 196 ValValPheValLys 200  
 DB 490 GTATCTTTAAAAAG 504

RESULT 8  
 AAX24296  
 ID AAX24296 standard; DNA; 519 BP.  
 XX

AC AAX24296;  
 XX 07-JUN-1999 (first entry)  
 DE Synthetic gene encoding winged bean trypsin inhibitor WTI-1B.  
 KW Trypsin inhibitor; winged bean; WTI-1B; chymotrypsin inhibitor;  
 KW WCI-3; protease inhibitor; enzyme engineering; protein engineering;  
 KW artificially synthesised gene; rice; insect resistance;  
 KW transgenic plant; ds.  
 XX Synthetic.  
 OS Psophocarpus tetragonolobus.  
 XX Key Location/Qualifiers  
 FH mutation 330 /\*tag= a  
 FT /\*note= "g to t mutation based on codon usage  
 FT of genes from Fabaceae plants"  
 FT mutation 348 /\*tag= b  
 FT /\*note= "t to c mutation, modifies ATTTA sequence  
 FT which causes mRNA instability"  
 FT mutation 381 /\*tag= c  
 FT /\*note= "g to a mutation based on codon usage  
 FT of genes from Fabaceae plants"  
 FT mutation 432 /\*tag= d  
 FT /\*note= "t to c mutation, modifies ATTTA sequence  
 FT which causes mRNA instability"  
 XX EP900842-A2.  
 PN 10-MAR-1999.  
 XX 01-SEP-1998; 98EP-0307004.  
 XX 01-SEP-1997; 97JP-0236332.  
 XX (NORQ ) TOHOKU NAT AGRIC EXPERIMENT STATION.  
 XX Mochizuki A;  
 XX WPI; 1999-155935/14.  
 XX P-PSDB; AAW97839.  
 PT Artificial synthetic trypsin inhibitor gene (WTI-1b) useful for  
 PT insect resistance in transgenic plants - designed on winged bean  
 PT chymotrypsin inhibitor (WCI-3) using codon and base comparisons  
 XX Disclosure; Page 19; 21pp; English.  
 XX This is the nucleotide sequence of an artificially synthesised gene  
 CC that codes for trypsin inhibitor WTI-1B (see AAW97839) of winged bean  
 CC (Psophocarpus tetragonolobus (L.) DC.). The synthetic gene is  
 CC designed for stable expression in rice plants. It was obtained by:  
 CC (i) comparing WTI-1B and winged bean chymotrypsin inhibitor WCI-3  
 CC (see AAW97840) amino acid sequences, and obtaining homologous amino  
 CC acid pairs (one each from WTI-1a and b); and (ii) selecting codons  
 CC from WCI-3 if the amino acid pair is the same, and from WTI-1B if  
 CC they are different. High usage codons are selected based on codon  
 CC usage of Fabaceae plants, and sequences which cause mRNA stability  
 CC are modified. A claimed synthetic gene (see AAX24282) includes DNA  
 CC encoding the WCI-3 signal peptide in addition to the sequence  
 CC encoding WTI-1B. Also claimed is a plant containing the claimed  
 CC synthetic gene. Transgenic plants can be produced that are  
 CC resistant to pests, especially lepidopteran insects.  
 XX Sequence 519 BP; 149 A; 112 C; 123 G; 135 T; 0 other;  
 SQ Alignment Scores:  
 Pred. No.: 3.94e-12 Length: 519

Score: 192.50 Matches: 53  
Percent Similarity: 50.81% Conservative: 41  
Best Local Similarity: 28.65% Mismatches: 66  
Query Match: 16.90% Indels: 25  
DB: 20 Gaps: 9

US-09-822-080B-2 (1-214) x AAX24296 (1-519)

QY 23 ProValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTyrtValLeuPro 41  
DB 4 CCAATGCTCGATTCGAAGGTAGTTCGAAATGGGCACATCTACTCTCTGCCA 63  
QY 42 LeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCys 61  
DB 64 GATAGTGGCA---CTCGGGGAGGANTAGACAGCAGCAGCAAGCAACATGTC 120  
QY 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSer 81  
DB 121 CCTCTAACAGTGGTACGATCTCCCAATGAGTCTCTGTAGGGAAACCATTAAGGATCTCA 180  
QY 82 AsnTrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAsp 101  
DB 181 TCC---CAATTGCGTTCAGGTTTCATCCCGATTACTCTCTAGTGCCTATTGGATTC--- 234  
QY 102 ValGluProThrIleCysAlaGlnSerAlaTyrtTrpValThrProAlaProSerPro 121  
DB 235 GCTAACCTCCCAAGTGT-----GCACCTCTCTCG 264  
QY 122 TrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGly---GlyGluAspSer 140  
DB 265 TGGTGGATGTTGTGAGGACCAACCAACCCCTCTGTAACCTAGTGAGCTAAAA 324  
QY 141 SerArgSerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrt 157  
DB 325 TCTACTAAATCGATATCTATTCAATTCGAGAAGTTACTCTAAGTTTCTCTCATAC 384  
QY 158 LysPheValPheCysSerGluGlyAsnAspCysIleAspValGly-----LysAsnGlu 175  
DB 385 AAGCTTAAGTACTGTGCGCAAGAGGACACCTGTAAGGATATCGGGATCTATAGGATCAG 444  
QY 176 GluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGlu 195  
DB 445 AAAGGATACGCACTTGTGTGCTGACTGACGAA-----AACCCATTAGTG 489  
QY 196 ValValPheValLys 200  
DB 490 GTATCTTTAAAAAG 504

RESULT 9

AAC33644

ID AAC33644 standard; DNA; 1092 BP.

XX AAC33644;

AC AAC33644;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 3803.

DE Arabidopsis thaliana.

XX Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.

```
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151085.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 23-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 4.59e-10
Score: 178.00
Percent Similarity: 45.33%
Best Local Similarity: 28.89%
Query Match: 15.63%
DB: 21

Length: 1092
Matches: 65
Conservative: 37
Mismatch: 87
Indels: 36
Gaps: 11

US-09-822-080b-2 (1-214) x AAC33644 (1-1092)

QY 4 MetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGly--- 22
Db 181 ATCTTCCTCTCTGCGCGCTTCATCAGCCACCGTGGAGTCCACCACCGAAGCAGCGTC 240
QY 23 ---ProValLeuAspThrAspGlyAspIlePheAspGly---SertYrValLeu 40
Db 241 GAACCAAGTTAAGGACATCAATGGAATACTCTCTTAACAGCGCTCAATACTACTACATCTG 300
QY 41 ProLeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60
Db 301 CCTGTAATTCGCGC---CGTGAAGCGGACTAACCATGCAAACTTAAAGACCGAAACC 357
QY 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80
Db 358 TGTCGCCACAACTGTGTCCAGACCAATTTGAAGTCTCTCAAGGCTTACCATGGATATTC 417
QY 81 SerAsnTrp---ArgSerArgValGlyPheValProGluGluGluAsnLeuAsnLys 99
Db 418 TCACCATATGACAAATCAAGAATC-----ATCCCTGTCTCAACTGATGTAAACATCAA 471
QY 100 MetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpValTrpProAlaPro 119
Db 472 TCC-----TCTCCAAT-----TCGATCTGGGAATTAGCAATTTTCGAC 510
QY 120 SerProTrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGluAsp 139
Db 511 GAGACGACGAAACAATGGTTTCATTTCACTTTGGGGTTCGAGGGAATCCCGGTGAGAAA 570
QY 140 SerSerArgSerPhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPhe 159
Db 571 ACGGTTGGTAACCTGTTCAAGATCGACAAATTTTGA-----AAAGATTACAAGATC 621
```

QY 160 ValPheCysSerGluGlyAsnAsp-----CysIleAspValGlyLysAsn 174  
 DB 622 AGCTTTGCTCTACTGCTGCAATTTCTGCAAGATCATCTGTAGAGAGCTGGAGTGT 681  
 QY 175 GluGluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPhe 194  
 DB 682 GTGCAAGATGGAAGAGAGACTTGTCTTAAGTGAT-----GTTCCACTA 726  
 QY 195 GluValValPheValLysAla-----ThrGlyThrAsp 205  
 DB 727 AAGGTATCTTCAAGAGCATCATACTAGATAGATACAAACAATTCACAAACACTCCA 786  
 QY 206 ThrSerSerLysThr 210  
 DB 787 ACGTCTCAGAGAGC 801

RESULT 10  
 AAX24283  
 ID AAX24283 standard; DNA; 624 BP.  
 XX  
 AC AAX24283;  
 XX  
 DT 07-JUN-1999 (first entry)  
 XX  
 DE Winged bean chymotrypsin inhibitor WCI-3 DNA.  
 XX  
 KW Chymotrypsin inhibitor; WCI-3; trypsin inhibitor; WTI-1B;  
 KW protease inhibitor; enzyme engineering; protein engineering;  
 KW artificially synthesised gene; winged bean; rice; transgenic plant;  
 KW insect resistance; ds.  
 XX  
 OS Psophocarpus tetragonolobus.  
 XX  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..72  
 FT /tag= a  
 FT mat\_peptide 73..621  
 FT /tag= b  
 XX  
 PN EP000842-A2.  
 XX  
 PD 10-MAR-1999.  
 XX  
 PF 01-SEP-1998; 98EP-0307004.  
 XX  
 PR 01-SEP-1997; 97JP-0236332.  
 XX  
 PA (NORQ ) TOHOKU NAT AGRIC EXPERIMENT STATION.  
 XX  
 PI Mochizuki A;  
 XX  
 DR WPI; 1999-155935/14.  
 DR P-PSDB; AAW97840.  
 XX  
 PT Artificial synthetic trypsin inhibitor gene (WTI-1b) useful for  
 PT insect resistance in transgenic plants - designed on winged bean  
 PT chymotrypsin inhibitor (WCI-3) using codon and base comparisons  
 XX  
 PS Disclosure; Page 14-15; 21pp; English.  
 XX  
 CC This DNA sequence encodes chymotrypsin inhibitor WCI-3 of winged  
 CC bean (Psophocarpus tetragonolobus (L.) DC.). An artificially  
 CC synthesised gene (see AAX24282) is claimed which codes for a trypsin  
 CC inhibitor (see AAW97839) designed for stable expression in rice.  
 CC The trypsin inhibitor comprises the winged bean WTI-1B polypeptide  
 CC having an N-terminal hydrophilic peptide derived from WCI-3 (i.e.  
 CC amino acid residues 1-24 of WCI-3) utilised for transporting the  
 CC protein precursor to a vacuole after translation. The synthetic  
 CC was obtained by: (i) comparing WTI-1B and WCI-3 amino acid  
 CC sequences, and obtaining homologous amino acid pairs (one each  
 CC from WTI-1a and b); and (ii) selecting codons from WCI-3 if the  
 CC amino acid pair is the same, and from WTI-1B if they are

CC different. High usage codons are selected and sequences which  
 CC cause mRNA stability are modified (see also AAX24294-96). Also  
 CC claimed is a plant containing the artificially synthesised gene.  
 CC Transgenic plants can be produced that are resistant to pests,  
 CC especially lepidopteran insects.  
 XX  
 SQ Sequence 624 BP; 178 A; 142 C; 134 G; 170 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 3 55e-09 Length: 624  
 Score: 167.00 Matches: 54  
 Percent Similarity: 46.30% Conservative: 46  
 Best Local Similarity: 25.00% Mismatches: 89  
 Query Match: 14.66% Indels: 28  
 DB: 20 Gaps: 10

US-09-822-080B-2 (1-214) x AAX24283 (1-624)

QY 3 ProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGly 22  
 DB 23 CCCTCTTTCTACTCTCTGCGCATCATCTCACACCTACCA-TCATCCACTGCTGAT---GAT 78  
 QY 23 ProValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTyrTyrValLeuPro 41  
 DB 79 GATTGTGCTGATGCTGAAGTAACTTGAATAATGGTGGCACATACTATCTGTTGCCA 138  
 QY 42 LeuIleTyrGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCys 61  
 DB 139 CATATATGGGCA---CACGGGGGAGGAAATAGAAACACGCAAAACAGAAACGACCATGC 195  
 QY 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSer 81  
 DB 196 CCTCTAACAGTGGTACGATCTCCCAATGAGTCTCTTAAGGGGAACCAATAAGGATCTCA 255  
 QY 82 AsnTrpArgSerArgValGlyPheValProGluGluAlaAsnLeuAsnIleLysMetAsp 101  
 DB 256 TCCCAATCTCTTTCATTTG---TTCCATCCCGAGAGGCTCTCTAGTGGCTCTTGGATTCC--- 309  
 QY 102 ValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerPro 121  
 DB 310 GCTAACCTCCATCTCTGCGAGCTTCTCCGTGGGACTGTTGTTGACTCTCCACAA--- 366  
 QY 122 TrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGlu----- 138  
 DB 367 -----GGACCGGCTGTTAACTTAGTCAGCAAAACTTCCG 402  
 QY 139 AspSerSerArgSerPhePheGlnIleLysLys---ThrGluAlaLysLeuAsnAlaTyr 157  
 DB 403 GAAAGGATATTCTAGTGTTTAAATTCGAGAAAGTTTCCCATTTCTAACATTCACGTGTAC 462  
 QY 158 LysPheValPheCysSerGluGlyAsnAspCysIleAspValGlyLys----- 173  
 DB 463 AAGCTTTGTACTCTCAACATGACGAGAGAGATGTGAAGTGTGATCATATATCGGGATT 522  
 QY 174 ---AsnGluGluGlyValArgValArgGlyLeuValLeuGlySerThrProPheAlaThr 192  
 DB 523 CATAGGATGCAATGGAACAGACAGCTTTGGTGGTACTGAG-----GAAAC 570  
 QY 193 ProPheGluValValPheValLysAlaThrGlyThrAspThrSerSer 208  
 DB 571 CCATTAGAGCTGTGCTCTTGAAAGCTAAGTCAGAAACTGCATCAAGC 618

RESULT 11  
 AAX24283  
 ID AAX24283 standard; DNA; 2352 BP.  
 XX  
 AC AAX24283;  
 XX  
 DT 08-JUL-1991 (first entry)  
 XX  
 DE gSPO-B1 insert encoding sporamin B.  
 XX  
 KW sporamin; promoter; potato; tuber; tissue-specific expression; ss.





XX Disclosure; Fig 4; 11pp; Japanese.  
 XX The gene product is sporamine, a soluble protein obtained from  
 CC Ipomoea betatas (sweet potato) it may be expressed from an  
 CC E.coli expression system transformed by the plasmid pimo336,  
 CC preferably with ampicillin resistance.  
 XX  
 SQ Sequence 954 BP; 258 A; 266 C; 213 G; 217 T; 0 other;

## Alignment Scores:

Pred. No.: 3.4e-08 Length: 954  
 Score: 160.50 Matches: 51  
 Percent Similarity: 44.51% Conservative: 30  
 Best Local Similarity: 28.02% Mismatches: 70  
 Query Match: 14.09% Indels: 31  
 DB: 7 Gaps: 9

US-09-822-080B-2 (1-214) x AAN60845 (1-954)

Qy 23 ProValLeuAspThrAspGlyAspIlePheAspGly---SerTyrTyrValLeuPro 41  
 Db 250 CCGGTCTGCATCATCAACGGCGACGAGTCCGGCGCGGAAATACATACATCTCTCC 309  
 Qy 42 LeuIleTyrProThrGlyGlyLeuThrLeuVal-----SerArgArgGlyAsn 59  
 Db 310 GCCATATGGGCGCGCGGAGGAGGCGCTGAGACTGCTCGATTGGATTCCTCTCGAAC 369  
 Qy 60 GlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79  
 Db 370 GAATGGCGCCAGCGACGTGATCTCCGGGAGCGACTTCGACAACGGCGACCGATTACC 429  
 Qy 80 PheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLys 99  
 Db 430 ATCAGC-----CCGGCGGACCGGAAATCCACCGTCGTC 462  
 Qy 100 MetAsp-----ValGluProThrIleCysAlaGln 109  
 Db 463 ATCGCGTCGAGCTCCAGACCTTCAGATTCAACATTGCGACCAACAACACTCTCGTAAAG 522  
 Qy 110 SerAlaTyrTrpTrpValThrProAlaProSerProTrpArgSerLeuPheIleAla 129  
 Db 523 AAGCTAAACTGGGGATC---AAGCAGCAGTGAATCCGGGCAATATTTCGTGAAGCC 579  
 Qy 130 GlyProLysProGluAlaGlyGlyGluAspSerSerArgSerPheGlnIleLysLys 149  
 Db 580 GGC-----GAGTCTGCTCCGACAATAGCAACAG-----TTCAGATTGAGGTG 624  
 Qy 150 ThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSerGluGlyAsnAsp---Cys 168  
 Db 625 GTCAACGACAACTTAACGGCTTACAAATCAGTTATTGTTCAGTTCGGCACCAGAAATGC 684  
 Qy 169 IleAspValGlyLysAsnGluGluGlyGlyValArgGly-----LeuValLeuGlySer 186  
 Db 685 TTCAACGCTGGCAGATACTACGACCCGCTTGACCAGGGCTACGCGTTTGCTCTCAGTAAT 744  
 Qy 187 ThrPro 188  
 Db 745 ACTCCT 750

## RESULT 13

AAZ07003  
 ID AAZ07003 standard; cDNA; 899 BP.

XX AAZ07003;

XX 09-NOV-1999 (first entry)

XX Sweet potato sporamin gene.

XX Sweet potato; sporamin; pest control; trypsin inhibitor;  
 KW insect resistance; tobacco cutworm; tobacco plant; ds.

XX

OS Ipomoea batatas.  
 PN EP945508-A1.  
 XX  
 PD 29-SEP-1999.  
 XX  
 PF 17-MAR-1998; 98EP-0301988.  
 XX  
 PR 17-MAR-1998; 98EP-0301988.  
 XX  
 PA (NASC-) NAT SCI COUNCIL.  
 XX  
 PI Chen Y, Kao S, Lin C, Lin M, Tuan S, Yeh K;  
 XX  
 DR WPI; 1999-530047/45.  
 XX  
 PT Use of sporamin gene for controlling insect pests on plants,  
 FT especially tobacco  
 XX  
 PS Claim 7; Fig 1; 22pp; English.  
 XX  
 CC The present sequence represents the sweet potato sporamin gene. The  
 CC sporamin polypeptide (I) can be used to control an insect pest on a  
 CC plant. A vector containing the sporamin gene can be transformed into  
 CC plant cells to form transgenic plant cells which may then be used to  
 CC cultivate first and second generation transgenic plants expressing (I).  
 CC (I) is a trypsin inhibitor and will control insect pests on plants  
 CC which express it, especially tobacco cutworms on tobacco plants. The  
 CC sporamin gene may be used to produce primers used to amplify sequences  
 CC related to the gene.  
 XX  
 SQ Sequence 899 BP; 228 A; 273 C; 189 G; 209 T; 0 other;

## Alignment Scores:

Pred. No.: 1.68e-06 Length: 899  
 Score: 145.00 Matches: 53  
 Percent Similarity: 41.79% Conservative: 31  
 Best Local Similarity: 26.37% Mismatches: 71  
 Query Match: 12.73% Indels: 46  
 DB: 20 Gaps: 9

US-09-822-080B-2 (1-214) x AAZ07003 (1-899)

Qy 23 ProValLeuAspThrAspGlyAsp---IleIlePheAspGlySerTyrTyrValLeuPro 41  
 Db 179 CCAGTACTGGACATCAACGGCGACGAGTCCGCGCGGGGAACTACTACATGTCTCC 238  
 Qy 42 LeuIleTyrProThrGlyGlyLeuThrLeuValSer---ArgArgGlyAsnGln 60  
 Db 239 GCCATATGGGAGCGCGCGGGGAGGGCTAAGACTCGCCCACTTGGACATGATGTCCAAA 298  
 Qy 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80  
 Db 299 TCGCGCAGCGAGTCATCGTATCCCCCAACGACTTAGACAACGGCGACCCCATCACCATC 358  
 Qy 81 SerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMet 100  
 Db 359 ACG-----CCGGCGGACGCGC 373  
 Qy 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrp----- 113  
 Db 374 GACCCGGAATCCACCGTGTCTATCGCTGACGACGTACACAGACTTCCCGGTTCACATCGCC 433  
 Qy 114 -----TrpValThrProAlaProSerProTrpArg 123  
 Db 434 ACCAACAAGCTCTCGTGAACAACGCTGAAGTGGGAATCCAGCAGCAGCGCGCTCCGGG 493  
 Qy 124 SerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerArgSer 143  
 Db 494 CAGTATTCTCTGAAAGCGCGC-----GAGTTGTGTCCGACAATAGCAACGAG--- 541  
 Qy 144 PhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSer 163  
 Db ----- 163



```
XX 27-MAR-1985; 85JP-0062887.
XX
XX 27-MAR-1985; 85JP-0062887.
XX
XX (MITS ) MITSUBISHI CORP.
XX (MITU ) MITSUBISHI CHEM IND KK.
XX
XX WPI: 1986-296104/45.
XX
XX P-PSDB; AAP61023.
XX
XX Prepn. of DNA segment and host - by extracting mRNA from Ipomoea
XX betatas, and converting to corresp. cDNA using vector-primer
XX method (by Okayama-Berg).
XX
XX Disclosure; Fig 3; 11pp; Japanese.
XX
XX The gene product is sporamine, a soluble protein obtained from
XX Ipomoea betatas (sweet potato) it may be expressed from an
XX E.coli expression system transformed by the plasmid pimo23,
XX preferably with ampicillin resistance.
XX
XX Sequence 894 BP; 230 A; 271 C; 184 G; 209 T; 0 other;
XX

Alignment Scores:
Pred. No.: 3.59e-06 Length: 894
Score: 142.00 Matches: 53
Percent Similarity: 41.29% Conservative: 30
Best Local Similarity: 26.37% Mismatches: 72
Query Match: 12.47% Indels: 46
DB: 7 Gaps: 9

US-09-822-080B-2 (1-214) x AAN60844 (1-894)
Qy 23 ProValLeuAspThrAspGlyAsp---IleIlePheAspGlySerTyrTyrValLeuPro 41
Db 185 CCGAGTACTCGACATCAACGGCGACGAGGTCCGCGCGGGGGAACCTACATGGTCTCC 244
Qy 42 LeuIleTrpGlyProThrGlyGlyLeuThrLeuValSer---ArgArgGlyAsnGln 60
Db 245 GCCATATGGGGGCGGGGGGGGGCTAAGACTCGCCCATCTGGACACGATGCCAAA 304
Qy 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyTleProValLysPhe 80
Db 305 TGGCGCCAGCGATCATCGTATCCGCCAACGACTTAGACAAGCGGACCCCATCACCATC 364
Qy 81 SerAsnTrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMet 100
Db 365 ACG-----CCGGCGACGGCC 379
Qy 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrp----- 113
Db 380 GACCCGAATCCACCGTGGTGCCTGCGCTACGACCTACGACACTTCGGTTCAACATCGCC 439
Qy 114 -----TrpValThrProAlaProSerProTrpArg 123
Db 440 ACCAACAACACTGTCGTGAACACAGTGAACCTGGGGAATCCAGCACGACGCGTCCGGG 499
Qy 124 SerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArgSer 143
Db 500 CAGTATTTCCTCAAGCCGCG-----GAGTTGTGTCGATTAATAGCAACCCAG--- 547
Qy 144 PhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSer 163
Db 548 ---TTCAAGATTGAGGTGTCGATGCCAACCTTAACCTTCTACAAACTCACTTACTGTCTAG 604
Qy 164 GluGlyAsnAsp---CysIleAspValGlyLysAsnGluGlyGlyValArgGlyLeu 182
Db 605 TTCGGTCCGACAAATGCTACAAACGTCGCGAGATTCACGACCCCATGTTGAGGACGACG 664
Qy 183 ValLeuGly---SerThrProProPheAlaThrProPheGluValValPheValLysAla 201
Db 665 CGTTTGGCTCTCTCCAAATCTCCCTTC-----GTTTTTGTCTCAACACCT 709
. . . .
```

Qy 202 Thr 202  
|||  
Db 710 ACC 712

Search completed: January 25, 2003, 18:31:34  
Job time : 268 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 25, 2003, 18:18:30 ; Search time 55 seconds  
(without alignments)  
1193.251 Million cell updates/sec

Title: US-09-822-080B-2  
Perfect score: 1139  
Sequence: 1 MNPMFYLLAFTTVAATAN.....EVFVKATGDTSSKTSII 214

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODEL=frame-p2n.model -DEV=xlh

-O=/cgn2.1/USPTO\_spool/US09822080/runat\_22012003\_142638\_4415/app\_query.fasta\_1.391  
-DB-Issued\_Patents\_NA -QMT=fastcap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09822080.qcgn.1.1.17 @runat\_22012003\_142638\_4415 -XCPU=6 -ICPU=3  
-NO\_XUPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:  
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/PCUTS.COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	222.5	19.5	917	1	US-07-949-812-1
2	200.5	17.6	591	1	US-09-143-211-4
3	200.5	17.6	603	4	US-09-143-211-15
4	200.5	17.6	603	4	US-09-143-211-16
5	200.5	17.6	603	4	US-09-143-211-17
6	194	17.0	516	2	US-08-943-814-1
7	194	17.0	539	2	US-08-702-703-1
8	194	17.0	541	2	US-08-943-814-9
9	167	14.7	624	4	US-09-143-211-2
10	145	12.7	899	4	US-09-038-542-1
11	94.5	8.3	2456	4	US-08-064-693A-19
12	94.5	8.3	6641	4	US-09-064-693A-25

13	90	7.9	111282	4	US-09-754-250-3	Sequence 3, Appli
14	88	7.7	9057	4	US-09-453-702B-194	Sequence 194, App
15	85.5	7.5	49136	4	US-09-422-869-1	Sequence 1, Appli
16	84.5	7.4	2559	4	US-09-118-408-43	Sequence 43, Appl
17	84.5	7.4	2559	4	US-09-506-855-43	Sequence 43, Appl
18	83.5	7.3	7808	2	US-08-149-097D-22	Sequence 22, Appl
19	83.5	7.3	7808	3	US-08-949-386-22	Sequence 22, Appl
20	83.5	7.3	7808	3	US-08-450-562-22	Sequence 22, Appl
21	83.5	7.3	7808	4	US-08-984-709A-22	Sequence 22, Appl
22	83.5	7.3	7808	4	US-08-450-272-22	Sequence 22, Appl
23	83	7.3	3563	4	US-09-041-886-20	Sequence 20, Appl
24	83	7.3	3596	2	US-08-779-801-5	Sequence 5, Appli
25	83	7.3	3596	4	US-09-298-441-3	Sequence 5, Appli
26	83	7.3	3632	2	US-08-779-801-3	Sequence 3, Appli
27	83	7.3	3632	4	US-09-298-441-3	Sequence 3, Appli
28	82.5	7.2	31571	1	US-08-323-443B-1	Sequence 1, Appli
29	82.5	7.2	53526	3	US-08-658-136-2	Sequence 2, Appli
30	82.5	7.2	53577	3	US-08-658-136-1	Sequence 1, Appli
31	82	7.2	2372	4	US-09-145-391-1	Sequence 1, Appli
32	81	7.1	3632	2	US-08-779-801-4	Sequence 4, Appli
33	81	7.1	3632	4	US-09-298-441-4	Sequence 4, Appli
34	81	7.1	4080	2	US-08-710-249-3	Sequence 3, Appli
35	81	7.1	4080	4	US-09-220-157A-3	Sequence 3, Appli
36	81	7.1	13875	2	US-08-734-344-1	Sequence 1, Appli
37	79.5	7.0	9960	3	US-08-822-586-46	Sequence 46, Appli
38	79.5	7.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
39	79	6.9	1235	1	US-08-095-726-13	Sequence 13, Appl
40	79	6.9	1235	1	US-08-095-726-15	Sequence 15, Appl
41	79	6.9	1235	1	US-08-096-623A-13	Sequence 13, Appl
42	79	6.9	1235	1	US-08-096-623A-15	Sequence 15, Appl
43	78.5	6.9	717	2	US-07-956-399-3	Sequence 3, Appli
44	77	6.8	1896	4	US-09-345-468-24	Sequence 24, Appl
45	77	6.8	1896	4	US-09-414-453A-24	Sequence 24, Appl

## ALIGNMENTS

## RESULT 1

US-07-949-812-1

; Sequence 1, Application US/07949812

; Patent No. 5668007

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: RECOMBINANT 21KD COCOA PROTEIN AND PRECURSOR

; NUMBER OF SEQUENCES: 28

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/949,812

; FILING DATE: 27-JAN-1993

; CLASSIFICATION: 435

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 917 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 21..686

; OTHER INFORMATION: /product= "DNA CODING SEQUENCE FOR

; OTHER INFORMATION: 21KD PROTEIN AND DEDUCED AMINO ACIDS"

US-07-949-812-1

Alignment Scores:

Pred. No.: 5.14e-17

Score: 222.50

Percent Similarity: 48.28%

Best Local Similarity: 35.47%

Length: 917

Matches: 72

Conservative: 26

Mismatches: 66



```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:artificially
; OTHER INFORMATION: synthesized wti-lb gene (mtwlib) designed based on
; OTHER INFORMATION: base sequence of WCI-3 gene
US-09-143-211-15

```

```

Alignment Scores:
Pred. No.: 1.09e-14 Length: 603
Score: 200.50 Matches: 57
Percent Similarity: 50.25% Conservative: 44
Best Local Similarity: 28.36% Mismatches: 75
Query Match: 17.60% Indels: 25
DB: 4 Gaps: 9

```

US-09-822-080B-2 (1-214) x US-09-143-211-15 (1-603)

```

QY 7 PheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGly?roValLeuAsp 26
DB 40 TTCTACTCTCTGCCATCATCTCACCTACCATCATCTGCTGAGCCATTGCTCGAT 99
QY 27 ThrAspGlyAspIleIlePheAsp---GlySerTyrTyrValLeuProLeuIleTrpGly 45
DB 100 TCTGAAGGTGAGTGTGCGAATGCTGCAATCATCTATCTGTGCCAGATAGATGGCA 159
QY 46 ProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCysProLeuPheIle 65
DB 160 ---CTCGGGGAGGAATAGACGACGACACAGGACCGAACATGCTCTTAACAGTG 216
QY 66 GlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSer 85
DB 217 GTACGATCTCCCAATGAGTCTCTGTAGGGGAACCATTAAGGATCTCATCC---CAATTG 273
QY 86 ArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspValGluProThr 105
DB 274 CGTTCAGGTTTCATCCCGGATCTCTAGTGGTATGGATTC---GCTAACCCCTCA 330
QY 106 IleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrpArgSerLeu 125
DB 331 AAGTGT-----GCACCTTCTCCGCTGGTGGACTGTT 360
QY 126 PheIleAlaAlaGlyProLysProGluAlaGly---GlyGluAspSerSerArgSerPhe 144
DB 361 GTTGAGGACCAACACACACCTCTGTAAACTTAGTGAGCTAAATCTACTAAATTC 420
QY 145 -----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPhe 161
DB 421 GATTATCTATTAAATTCGAGAAAGTTACCTTAAGTTTCTCTCATACAGTTAAGTAC 480
QY 162 CysSerGluGlyAsnAspCysIleAspValGly-----LysAsnGluGluGlyVal 179
DB 481 TGTGCCAAGAGGGACACCTGTAAGGATATCGGGATTTATAGGATCAGAAAGGATACGCA 540
QY 180 ArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGluValValPheVal 199
DB 541 CGTTGGTGGTACTGACGAA-----AACCCATTAGTGGTATCTTTAA 585
QY 200 Lys 200
DB 586 AAG 588

```

#### RESULT 4

```

US-09-143-211-16
; Sequence 16, Application US/09143211B
; Patent No. 6310275
; GENERAL INFORMATION:
; APPLICANT: Mochizuki, Atsushi
; APPLICANT: Director General of Tohoku National Agricultural Experiment
; APPLICANT: Station, Ministry of Agriculture, Forestry and Fisheries
; TITLE OF INVENTION: Artificially Synthesized Gene for Trypsin Inhibitor
; FILE REFERENCE: 085761-00010005
; CURRENT APPLICATION NUMBER: US/09/143,211B

```

```

; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: JP 9-236332
; EARLIER FILING DATE: 1997-09-01
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:artificially
; OTHER INFORMATION: synthesized wti-lb gene (mtwlib) obtained by
; OTHER INFORMATION: modifying the sequence shown in Figure 1
; OTHER INFORMATION: (SEQ ID NO:18) based on codon usage of genes
; OTHER INFORMATION: from Fabaceae plant
US-09-143-211-16

```

```

Alignment Scores:
Pred. No.: 1.09e-14 Length: 603
Score: 200.50 Matches: 57
Percent Similarity: 50.25% Conservative: 44
Best Local Similarity: 28.36% Mismatches: 75
Query Match: 17.60% Indels: 25
DB: 4 Gaps: 9

```

US-09-822-080B-2 (1-214) x US-09-143-211-16 (1-603)

```

QY 7 PheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGlyProValLeuAsp 26
DB 40 TTCTACTCTCTGCCATCATCTCACCTACCATCATCTGCTGAGCCATTGCTCGAT 99
QY 27 ThrAspGlyAspIleIlePheAsp---GlySerTyrTyrValLeuProLeuIleTrpGly 45
DB 100 TCTGAAGGTGAGTGTGCGAATGCTGCAATCATCTATCTGTGCCAGATAGATGGCA 159
QY 46 ProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCysProLeuPheIle 65
DB 160 ---CTCGGGGAGGAATAGACGACGACACAGGACCGAACATGCTCTTAACAGTG 216
QY 66 GlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSer 85
DB 217 GTACGATCTCCCAATGAGTCTCTGTAGGGGAACCATTAAGGATCTCATCC---CAATTG 273
QY 86 ArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspValGluProThr 105
DB 274 CGTTCAGGTTTCATCCCGGATCTCTAGTGGTATGGATTC---GCTAACCCCTCA 330
QY 106 IleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrpArgSerLeu 125
DB 331 AAGTGT-----GCACCTTCTCCGCTGGTGGACTGTT 360
QY 126 PheIleAlaAlaGlyProLysProGluAlaGly---GlyGluAspSerSerArgSerPhe 144
DB 361 GTTGAGGACCAACACACACCTCTGTAAACTTAGTGAGCTAAATCTACTAAATTC 420
QY 145 -----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPhe 161
DB 421 GATTATCTATTAAATTCGAGAAAGTTACCTTAAGTTTCTCTCATACAGTTAAGTAC 480
QY 162 CysSerGluGlyAsnAspCysIleAspValGly-----LysAsnGluGluGlyVal 179
DB 481 TGTGCCAAGAGGGACACCTGTAAGGATATCGGGATTTATAGGATCAGAAAGGATACGCA 540
QY 180 ArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGluValValPheVal 199
DB 541 CGTTGGTGGTACTGACGAA-----AACCCATTAGTGGTATCTTTAA 585
QY 200 Lys 200
DB 586 AAG 588

```

#### RESULT 5

US-09-143-211-17





Db 4 TTATTAGATGTTAAACGGCGAAAGTGTGTCAGAACGGCGGTACCTATTATCTGCTGCCGAG 63  
QY 43 IleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysPro 62  
Db 64 GTGTGGCG---CAGGGCGCGCGGTGCGACGTGGCGAAACCGCGGAGAAACCTGCCCG 120  
QY 63 LeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82  
Db 121 CTGACCGTGGTCAGAGCCGACGAACGACGAGCGATGGCAACCGATCGTATTGAAGC 180  
QY 83 TrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAspVal 102  
Db 181 ---CGTCTCGTAGCGCGTTTATTCGGATGATGATAAAGTGGCTATTGGCTTTCGCTAT 237  
QY 103 GluProThrIleCysAlaGlnSerAlaTrpTrpValThrProAlaProSerProTrp 122  
Db 238 GCGCCGAAA---TGC CGCGCGAGCCGCGTGGTGGACCGTGGAGATGAACAGGAAGC 294  
QY 123 ArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArg 142  
Db 295 CTGACGCTGAACCTGAGC-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTrpLysPhe 159  
QY 143 SerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTrpLysPhe 159  
Db 328 CAGTTTGATATCCGTTTAAATTTGAACAGGTGAGCGATGACGTGATGCTATAAAGT 387  
QY 160 ValPheCysSerGluGly-----AsnAspCysIleAspValGlyLysAsn---GluGlu 176  
Db 388 CTGTATTGC---GAAGGCAACATGAAATAATGCGGAGCATTTGGCATTAACGCTGATCAG 444  
QY 177 GlyGlyValArgGlyLeuValLeuGlySerThrProPropheAlaThrProPheGluVal 196  
Db 445 AAAGGCTATCGTCTGTGGTGTG-----ACCGAAGATTAT-----CCGCTGACCGCTG 492  
QY 197 ValPheValLysAlaThrClyThrAspThrSerSer 208  
Db 493 GTGCTGAAAAA-----GATGAAGACGAGC 516

RESULT 7  
US-08-702-703-1  
; Sequence 1, Application US/08702703  
; Patent No. 5958722  
; GENERAL INFORMATION:  
; APPLICANT: KOHNERT, Ulrich  
; APPLICANT: STERN, Anne  
; APPLICANT: FISCHER, Stephan  
; TITLE OF INVENTION: USE OF A RECOMBINANT INHIBITOR FROM  
; TITLE OF INVENTION: ERYTHRINA CAFFRA FOR PURIFYING SERINE PROTEASES  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nikaido Marmelstein Murray and Oram LLP  
; STREET: 655 Fifteenth St. N.W., Suite 330  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/702,703  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP95/00926  
; FILING DATE: 13-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P4408939.2  
; FILING DATE: 16-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 4424171.2

; FILING DATE: 8-JUL-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitts, Monica C.  
; REGISTRATION NUMBER: 36,105  
; REFERENCE/DOCKET NUMBER: 1614-6046  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 638-5000  
; TELEFAX: (202) 638-4810  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 539 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 9..527  
; OTHER INFORMATION: /note= "Met only included in  
; OTHER INFORMATION: prokaryotic expression"  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..8  
; OTHER INFORMATION: /function= "multiple cloning site"  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 528..539  
; OTHER INFORMATION: /function= "multiple cloning site"  
; US-08-702-703-1  
Alignment Scores:  
Pred. No.: 5,39e-14 Length: 539  
Score: 194.00 Matches: 58  
Percent Similarity: 50.52% Conservative: 39  
Best Local Similarity: 30.21% Mismatches: 67  
Query Match: 17.03% Indels: 28  
DB: 2 Gaps: 12  
US-09-822-080B-2 (1-214) x US-08-702-703-1 (1-539)  
QY 24 ValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTrpTrpValLeuProLeu 42  
Db 15 TTATTAGATGTTAAACGGCGAAAGTGTGTCAGAACGGCGGTACCTATTATCTGCTGCCGAG 74  
QY 43 IleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysPro 62  
Db 75 GTGTGGCG---CAGGGCGCGCGTGTGTCAGCTGCGGCGGCGGCGGCGGCGGCGGCGG 131  
QY 63 LeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsn 82  
Db 132 CTGACCGTGGTGCAGACCGCGACGAGACTGAGCGATGGCAACCGATCGTATTGAAGC 191  
QY 83 TrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAspVal 102  
Db 192 ---CGTCTGCTAGCGCGTTTATTCGGATGATGATAAAGTGGCTATTGGCTTTCGCTAT 248  
QY 103 GluProThrIleCysAlaGlnSerAlaTrpTrpValThrProAlaProSerProTrp 122  
Db 249 GCGCCGAAA---TGC CGCGCGAGCCGCGTGGTGGACCGTGGTGGAGATGAACAGGAAGC 305  
QY 123 ArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArg 142  
Db 306 CTGACGCTGAACCTGAGC-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTrpLysPhe 159  
QY 143 SerPhe-----PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTrpLysPhe 159  
Db 339 CAGTTTGATATCCGTTTAAATTTGAACAGGTGAGCGATGACGTGATGCTATAAAGT 398  
QY 160 ValPheCysSerGluGly-----AsnAspCysIleAspValGlyLysAsn---GluGlu 176  
Db 399 CTGTATTGC---GAAGGCAACATGAAATAATGCGGAGCATTTGGCATTAACGCTGATCAG 455  
QY 177 GlyGlyValArgGlyLeuValLeuGlySerThrProPropheAlaThrProPheGluVal 196



Best Local Similarity: 25.00% Mismatches: 89  
 Query Match: 14.66% Indels: 28  
 DB: 4 Gaps: 10

US-09-822-080B-2 (1-214) x US-09-143-211-2 (1-624)

QY 3 ProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGly 22  
 DB 23 CCCTCTTCTACTCTCTGCATCTCTACACCTACCA-TCATCCACTGCTGAT--GAT 78  
 QY 23 ProValLeuAspThrAspGlyPhePheAsp---GlySerTyrTyrValLeuPro 41  
 DB 79 GATTGGTCGATCGCTGAGGTAACTTGTGAATGTGGCACAATATCTGTGTGCA 138  
 QY 42 LeuileTrpGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGlnCys 61  
 DB 139 CATATATGGCA---CACGGGGAGGAATAGAAACAGCAAAACAGGAACGACCATGC 195  
 QY 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSer 81  
 DB 196 CCTCTAACAGTGTGATCTCTCCCAATGAGGTCTCTAAAGGGAACCAATAAGGATCTCA 255  
 QY 82 AsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAsp 101  
 DB 256 TCCCAATCCITTCATG---TTCATCCCAAGAGCTCTCTAGTGGCTCTTGGATTC--- 309  
 QY 102 ValGluProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSerPro 121  
 DB 310 GCTAACCCCTCCATCTCTGCGAGCTTCTCCGTTGGTGTGTGTGATCTCTCCACA--- 366  
 QY 122 TrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGlu----- 138  
 DB 367 -----GGACCCGCTGTAAACTTAGTCAGCAAAACTTCCG 402  
 QY 139 AspSerSerArgSerPheGlnIleLysLys---ThrGluAlaLysLeuAsnAlaTyr 157  
 DB 403 GAAAGGATATCTAGTGTAAATTCAGAAAGATTCCCAATCTTAACATTCAGGTGTAC 462  
 QY 158 LysPheValPheCysSerGluGlyAsnAspCysIleAspValGlyLys----- 173  
 DB 463 AAGCTTTTGTACTCTCAACATGACGAGAGGATGTGAAGTGTGATCATATATCGGATT 522  
 QY 174 ---AsnGluGluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThr 192  
 DB 523 CATAGGATCGCATGGAACAGACAGCTTTGTGTGCTGACTGAG-----GAAAC 570  
 QY 193 ProPheGluValValPheValLysAlaThrGlyThrAspThrSerSer 208  
 DB 571 CCATTAGAGCTGTCTCTCAAGACTAAGTCAGAAACTGCATCAAGC 618

## RESULT 10

US-09-038-542-1  
 ; Sequence 1, Application US/09038542  
 ; Patent No. 6297427

## GENERAL INFORMATION:

APPLICANT: YEH, Kai-Wun  
 APPLICANT: LIN, Mei-In  
 APPLICANT: TUAN, Shu-Jen  
 APPLICANT: CHEN, Yih-Ming  
 APPLICANT: LIN, Chu-Yung  
 APPLICANT: KAO, Suey-Sheng  
 TITLE OF INVENTION: The insect-resistant use of sweet potato  
 TITLE OF INVENTION: Sporamin gene and method for controlling pests using the  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: McDonnell Boehnen Hulbert & Berghoff  
 STREET: 300 S. Wacker Drive, Suite 3200  
 CITY: Chicago  
 STATE: IL  
 COUNTRY: USA  
 ZIP: 60606  
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/038,542  
 FILING DATE: 03/11/98  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: TW 86103072  
 FILING DATE: 11-MAR-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CHAO, Mark  
 REGISTRATION NUMBER: 37,293  
 REFERENCE/DOCKET NUMBER: 98,250  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312-913-0001  
 TELEFAX: 312-913-0002  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 899 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 56..712  
 NAME/KEY: polyA\_site  
 LOCATION: one-of(796, 842, 856)  
 US-09-038-542-1

Alignment Scores:  
 Pred. No.: 6.97e-08 Length: 899  
 Score: 145.00 Matches: 53  
 Percent Similarity: 41.79% Conservative: 31  
 Best Local Similarity: 26.37% Mismatches: 71  
 Query Match: 12.73% Indels: 46  
 DB: Gaps: 9

US-09-822-080B-2 (1-214) x US-09-038-542-1 (1-899)

QY 23 ProValLeuAspThrAspGlyAsp---IleIlePheAspGlySerTyrTrpValLeuPro 41  
 DB 179 CCAGTACTGGACATCAACGGCGAGAGTCCCGCGCGGGGAACACTACTACATGGTCTCC 238  
 QY 42 LeuileTrpGlyProThrGlyGlyLeuThrLeuValSer---ArgArgGlyAsnGln 60  
 DB 239 GCCATATGGGAGCGCGGGGAGGGCTAAGACTCGCCCTTGGACATGATGTCCAAA 298  
 QY 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80  
 DB 299 TGGCCGACGACGTCATCGTATATCCCAACAGACTTAGACAACGGCGACCCCATCCATC 358  
 QY 81 SerAsnTrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMet 100  
 DB 359 ACG-----CCGGGAGGGCC 373  
 QY 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrp----- 113  
 DB 374 GACCCGGAATCCACCGTGTGCTGCGCTACGACGTTCCGGTTCAACATCGCC 433  
 QY 114 -----TrpValThrProAlaProSerProTrpArg 123  
 DB 434 ACCAACAGCTCTGCTGTAACACGCTGAAGTGGGAATCCAGCAGACAGCCGCTCCGGG 493  
 QY 124 SerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArgSer 143  
 DB 494 CAGTATTCTCTGAAAGCCGCG-----GAGTTGTGTCGCAACATAGCAACAG--- 541  
 QY 144 PhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSer 163  
 DB 144 PhePheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSer 163

Db 542 ---TTCAAGATTGAGTGGTGGATGCCAACCTTAACCTCTACAAACTCACTTACTGTCTGAG 598  
Qy 164 GluGlyAsnAsp---CysIleAspValGlyLysAsnGluGluGlyValArgGlyLeu 182  
Db 599 TTCGGCTCCGATAAATGCTACAACTGCGCAGATTCCACGACCAACATGTTGAGGACCCAG 658  
Qy 183 ValLeuGly---SerThrProPheAlaThrProPheGluValValPheValLysAla 201  
Db 659 CGTTTGGCTCTCTCCAAATCTCCCTTC-----GTTTTTGTCTATCAAACT 703  
Qy 202 Thr 202  
Db 704 ACC 706

## RESULT 11

US-09-064-693A-19/c  
; Sequence 19, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; OF A SPECIFIC PLASTICS PRECURSOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEEL--Lockheed Martin Idaho  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: Toshiba Satellite Pro T2150CDS  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word Perfect 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/064,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 22,387  
; REFERENCE/DOCKET NUMBER: LIT-PI-296  
; TELEPHONE: (208)526-9469  
; TELEFAX: (208)526-8339  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2456 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-064-693A-19

Alignment Scores:  
Pred. No.: 0.282 Length: 2456  
Score: 94.50 Matches: 42  
Percent Similarity: 40.00% Conservative: 20  
Best Local Similarity: 27.10% Mismatches: 72  
Query Match: 8.30% Indels: 21  
DB: 4 Gaps: 6

US-09-822-080b-2 (1-214) x US-09-064-693A-19 (1-2456)

Qy 51 LeuThrLeuValSerArgGlyAsnGlnCysProLeuPheIleGlyGlnGluArgSer 70  
Db 1608 TTGATGATTAATTCACAGCGCGAGGCCCGCCCTTCGTTAGGACGGCGAAGTCG 1549

Qy 71 GluValAsnArgGlyIleProValLysPheSerAsnTrpArgSerArgValGlyPheVal 90  
Db 1548 AAACACCGGTCCGGGTACCGAGCTTCATCATCATCTTCAACGCGAGATCCGCGATCGTT 1489  
Qy 91 ProGluGluGluAsnLeuAsnIleLysMetAsp-----ValGlu 103  
Db 1488 TCTACGCTGGAGCTTTTCACGCTAAATTCAGCAACTCCGCATAGACGTAACCGGTATCA 1429  
Qy 104 ProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrpArg 123  
Db 1428 CCT-----TCGGCACAAGAAACAGTGTCTCACCGTCTTTTCATCCGT 1384  
Qy 124 SerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSerArgSer 143  
Db 1383 TCTGTT-----GCATCTCCACAGCCCACTACCCCGGAATGCCAGTTCACGAGCG 1333  
Qy 144 PhePhe-----GlnIleLys-----LysThrGluAlaLysLeuAsnAlaTyrLys 158  
Db 1332 ATGATCGCGCGTGACAGGTACGACGCCGCGGTTGGTGACGATGGCAGATGCTTTC--- 1276  
Qy 159 PheValPheCysSerGluGlyAsnAspCysIleAspValGlyLysAsnGluGluGly 178  
Db 1275 TTCATGATCGGTTCCAGTCCGCGTCCGTCATGTCAGTAACACGACGTCGCCAGGTTCG 1216  
Qy 179 ValArgGlyLeuValLeuGlySerThrProPheAlaThrPro 193  
Db 1215 ATGCGGTTTCATTCGCTGATGTCATGATGACCTTCACCGGACCC 1171

## RESULT 12

US-09-064-693A-25/c  
; Sequence 25, Application US/09064693A  
; Patent No. 6210937  
; GENERAL INFORMATION:  
; APPLICANT: Ward, Thomas E.  
; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY  
; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION  
; OF A SPECIFIC PLASTICS PRECURSOR  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: W. Gary Goodson  
; ADDRESSEE: INEEL--Lockheed Martin Idaho  
; STREET: P.O. Box 1625  
; CITY: Idaho Falls  
; STATE: Idaho  
; COUNTRY: USA  
; ZIP: 83415-3810  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
; COMPUTER: Toshiba Satellite Pro T2150CDS  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: Word Perfect 7.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/064,693A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: W. Gary Goodson  
; REGISTRATION NUMBER: 22,387  
; REFERENCE/DOCKET NUMBER: LIT-PI-296  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (208)526-9469  
; TELEFAX: (208)526-8339  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6641 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

## US-09-064-693A-25

## Alignment Scores:

Pred. No.: 1.23 Length: 6641  
 Score: 94.50 Matches: 42  
 Percent Similarity: 40.00% Conservative: 20  
 Best Local Similarity: 27.10% Mismatches: 72  
 Query Match: 8.30% Indels: 21  
 DB: 4 Gaps: 6

US-09-822-080B-2 (1-214) x US-09-064-693A-25 (1-6641)

QY 51 LeuThrLeuValSerArgArgGlyAsnGlnCysProLeuPheIleGlyGlnLeuArgSer 70  
 DB 5793 TTGATGATAAATCCAGAGCCCAAGGCCACGCTTCGTTGGTAGCAGCGGAAGTCG 5734  
 QY 71 GluValAsnArgGlyIleProValLysPheSerAsnTrpArgSerArgValGlyPheVal 90  
 DB 5733 AAAGCAGCGTCGCGGTACCGACGTTTCATCATCTTCAACGGCAGATCCGGCATCGTT 5674  
 QY 91 ProGluGluGluAsnLeuAsnIleLysMetAsp-----ValGlu 103  
 DB 5673 TCTACGCTGGAGCTTTTCAGCGTAATTCAGCAACTCCGCATAGACGTACCGGTATCA 5614  
 QY 104 ProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSerProTrpArg 123  
 DB 5613 CTT-----TCGGCACAAAGAAACAGTACGTTCTCACCGTCTTTTCATCGT 5569  
 QY 124 SerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerArgSer 143  
 DB 5568 TCTGTT-----GCATCTCCACAGCCACTACCGCCGGAATGCCAGTTCACGAGCG 5518  
 QY 144 PhePhe-----GlnIleLys-----LysThrGluAlaLysLeuAsnAlaTyrLys 158  
 DB 5517 ATGATCGCGCGTGACAGGTACAGCGACCGCGTGTGTGACGATGGCAGATGCTTC- 5461  
 QY 159 PheValPheCysSerGluGlyAsnAspCysTyleaspValGlyLysAsnGluGluGly 178  
 DB 5460 TTCATGATCGGTCCAGTCCGCGTGGTTCATGTCAGTAACCAACGACGTCGCCAGGTCG 5401  
 QY 179 ValArgGlyLeuValLeuGlySerThrProPheAlaThrPro 193  
 DB 5400 ATGCGGTTTCATTCGCTCATGTCATGATGATGACTTTCACCGGACCC 5356

## RESULT 13

US-09-754-250-3  
 ; Sequence 3, Application US/09754250  
 ; Patent No. 6376225  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEI, Ming-Hui et al  
 ; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE  
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
 ; TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF  
 ; FILE REFERENCE: CL001063  
 ; CURRENT APPLICATION NUMBER: US/09/754,250  
 ; CURRENT FILING DATE: 2001-01-05  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 111282  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)...(111282)  
 ; OTHER INFORMATION: n = A,T,C or G  
 ; US-09-754-250-3

## Alignment Scores:

Pred. No.: 274 Length: 111282  
 Score: 90.00 Matches: 48  
 Percent Similarity: 37.77% Conservative: 23  
 Best Local Similarity: 25.53% Mismatches: 79

Query Match: 7.90% Indels: 38  
 DB: 4 Gaps: 10  
 US-09-822-080B-2 (1-214) x US-09-754-250-3 (1-111282)  
 QY 45 GlyProThrGlyGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysProLeuPhe 64  
 DB 65761 GGAGCCCGAGTGGGAGGTCAGACTTGCCTGGAGAGAGAACACACACTCTCCCTG 65820  
 QY 65 IleGlyGlnGluArgSerGluValAsnArgGlyIle-----Pro 77  
 DB 65821 GAGGGGATCCAGAGAGGAGATCACTTCATTCATTCGTCATTCATCCACCC 65880  
 QY 78 ValLysPheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsn 97  
 DB 65881 ATTCAATATTCTTGGCCATCA-----TTTCTGAGGATGTAAA 65922  
 QY 98 IleLysMetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrPro 117  
 DB 65923 CTCTCTTCTGACACTGACCCAGCGGACACTCAGGTCCTCC-----TCCTCTCTCT 65973  
 QY 118 Ala-----ProSerProTrpArgSerLeu---PheIleAlaAlaGlyProLys 132  
 DB 65974 GCTTGAGCCACCATGCTCTTCTTGAGGCTCTCTGAGGCTCTGCTCAGCTCCCAAC 66033  
 QY 133 ProGluAlaGlyGlyGlu---AspSerSerArgSerPhePheGlnIleLysThrGlu 151  
 DB 66034 CCACCTTGAGGGGTGAGGCTGAGGAGGTGTACAGACATTCAGGTCACCAACTCAGA 66093  
 QY 152 Ala---LysLeuAsnAlaTyrLysPheValPheCysSerGluGlyAsnAspCysIleAsp 170  
 DB 66094 GCTGGAGGCTGCCACTCACCAGGGGCTTCTCA-----66129  
 QY 171 ValGlyLysAsnGluGlyGlyValArgGlyLeuValLeuGlySerThrPro 188  
 DB 66130 ---GGCACAGGCTCCCTGCTGGTGGC---AGGGCTTGGCCCTTGTGACACCCCTTGGGG 66183  
 QY 189 -----ProPheAlaThrProPheGluValValPheValLysAlaThrGlyThrAsp 205  
 DB 66184 ACTAGAGCCCTCATCTCATCTCTGCTCAGGCTCTCTTTTGTGGCGGAGCTCTGATTAC 66243  
 QY 206 ThrSerSerLysThrMetSerIle 213  
 DB 66244 AGTGCCCAAAATCTGCCTCTTG 66267

RESULT 14  
 US-09-453-702B-194  
 ; Sequence 194, Application US/09453702B  
 ; Patent No. 6365723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blattner, Frederick R.  
 ; Burland, Valerie  
 ; Perna, Nicole T.  
 ; Plunkett, Guy  
 ; Welch, Rod  
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: 1 South Pinckney Street  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: US  
 ; ZIP: 53701-2113  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/453,702B  
 ; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <Unknown>



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 25, 2003, 18:19:35 ; Search time 60 Seconds  
(without alignments)  
1602.378 Million cell updates/sec

Title: US-09-822-080B-2

Perfect score: 1139

Sequence: 1 MNPMYFLLAFTTVLAATAN.....EVFVKATGTDTSKTSII 214

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cgn2.1/USPTO.spool/US09822080/runat\_22012003\_142639\_4427/app\_query.fasta\_1.391  
-DB=PublishedApplications\_NA -QFW=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=BLOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USRP=US09822080@cgn\_1.1.24 -runat\_22012003\_142639\_4427  
-NCPU=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA.\*  
1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2.6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2.6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2.6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	506.5	44.5	792	10	US-09-770-445-843 Sequence 843, App
2	173	15.2	412	10	US-09-878-574-3400 Sequence 3400, App
3	160.5	14.1	630	9	US-09-938-842A-1465 Sequence 1465, App
4	160	14.0	275	10	US-09-878-574-15072 Sequence 15072, A

#### ALIGNMENTS

##### RESULT 1

US-09-770-445-843

; Sequence 843, Application US/09770445

; Patent No. US20020023281A1

; GENERAL INFORMATION:

; APPLICANT: Gorlach, Jorn

; APPLICANT: An, Yong-Qiang

; APPLICANT: Hamilton, Carol M.

; APPLICANT: Price, Jennifer L.

; APPLICANT: Raines, Tracy M.

; APPLICANT: Yu, Yang

; APPLICANT: Rameaka, Joshua G.

; APPLICANT: Page, Amy

; APPLICANT: Matthew, Abraham V.

; APPLICANT: Ledford, Brooke L.

; APPLICANT: Woessner, Jeffrey P.

; APPLICANT: Haas, William David

; APPLICANT: Garcia, Carlos A.

; APPLICANT: Krickner, Maja

; APPLICANT: Slader, Ted

; APPLICANT: Davis, Keith R.

; APPLICANT: Allen, Keith

; APPLICANT: Hoffman, Neil

; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; TITLE OF INVENTION: Thailand

; FILE REFERENCE: 2023US (PARA-012PRV)

; CURRENT APPLICATION NUMBER: US/09/770,445

Sequence 3061, Ap  
Sequence 13194, A  
Sequence 14572, A  
Sequence 14101, A  
Sequence 11210, A  
Sequence 11929, A  
Sequence 7935, Ap  
Sequence 13721, A  
Sequence 3353, Ap  
Sequence 11909, A  
Sequence 69, Appl  
Sequence 2609, Ap  
Sequence 2231, Ap  
Sequence 5564, Ap  
Sequence 13069, A  
Sequence 14953, A  
Sequence 15595, A  
Sequence 1557, Ap  
Sequence 9051, Ap  
Sequence 7385, Ap  
Sequence 3950, Ap  
Sequence 12518, A  
Sequence 14258, A  
Sequence 2820, Ap  
Sequence 9421, Ap  
Sequence 15425, A  
Sequence 1946, Ap  
Sequence 922, App  
Sequence 11801, A  
Sequence 3, Appl1  
Sequence 14712, A  
Sequence 5011, Ap  
Sequence 13183, A  
Sequence 13242, A  
Sequence 1, Appl1  
Sequence 11239, A  
Sequence 263, App  
Sequence 43, Appl  
Sequence 699, App  
Sequence 3689, Ap  
Sequence 3, Appl1

; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 60/178,472  
; PRIOR FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 999  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 843  
; LENGTH: 792  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-770-445-843

Alignment Scores:  
Pred. No.: 1,56e-50 Length: 792  
Score: 506.50 Matches: 117  
Percent Similarity: 63.59% Conservative: 21  
Best Local Similarity: 53.92% Mismatches: 64  
Query Match: 44.47% Indels: 15  
DB: 10 Gaps: 7

US-09-822-080b-2 (1-214) x US-09-770-445-843 (1-792)

Qy 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrThrValLeuAlaAlaThrAlaAsn 20  
Db 96 ATGAATCCTAAGTTTACTTAGTTCTTGCTTAACCCGCGTCTCGCCCTCAACGCA--- 152  
Qy 21 AlaGlyProValLeuAspThrAspGlyAspIlePheAspGlySerTyrTrpValLeu 40  
Db 153 TATGGTGGCGTTGTAGACATCGATGAAACACCATGTTCCAGAAAGTTACTACGTTCTC 212  
Qy 41 ProLeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60  
Db 213 CCGTGCATCCGTGGC---CGAGCGGAGGCGCTGACCTAGACAGCGCGGTGGCAGCCA 269  
Qy 61 CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80  
Db 270 TGTCTTACCATATGCTGAGGAATCTTCAGAAAGTTGATGAGGCGATTCCTCGTAANAATTC 329  
Qy 81 SerAsnTrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMet 100  
Db 330 TCAAACTGGAGGCTTAAGTTGGTTCGTTCCGGAATCAGAACCTCAACATCGAAACA 389  
Qy 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSer 120  
Db 390 GAGCTGGAGCCAGCATCTGCATCCAGTCAACCTACTGCGGGTCTGGTGGATTTGACCAC 449  
Qy 121 ProTrpArgSerLeuPheIleAlaGlyProLysProGluAlaGlyGlyGluAspSer 140  
Db 450 GAGAGGAGCAGTACTCTGCTGGTGTGTCCTCAAAACCCAGAGGGTTCGGAACAAGATTCC 509  
Qy 141 SerArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheVal 160  
Db 510 TTGAAGAGATTCTTCAAGATCGAGAAATCTGGAGAG-----GATGCTTACAAGTTTGTG 563  
Qy 161 Phe-----CysSerGluGlyAsn---AspCysIleAspValGly---LysAsn 174  
Db 564 TTCTGTCTCCGACTTCGCACTTCGCAATCCAAATCGAGCGATCTCGGGATATTATCAT 623  
Qy 175 GluGluGlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPhe 194  
Db 624 GATGAATTGGCGTTCGTTGGCTTTAAGCGAT-----AAGCGGTTTC 668  
Qy 195 GluValValPheValLysAlaThrGlyThrAspThrSerSerLysThrMet 211  
Db 669 TTGTTATGTTCAAAAAGCAATATGTGACCGAAGTTTCGTCCAAAGACTATG 719

## RESULT 2

US-09-878-574-3400  
; Sequence 3400, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 3400  
; LENGTH: 412  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: LIB3028-011-Q1-B1-F6  
US-09-878-574-3400

Alignment Scores:  
Pred. No.: 9.64e-12 Length: 412  
Score: 173.00 Matches: 46  
Percent Similarity: 60.98% Conservative: 29  
Best Local Similarity: 37.40% Mismatches: 38  
Query Match: 15.19% Indels: 10  
DB: 10 Gaps: 7

US-09-822-080b-2 (1-214) x US-09-878-574-3400 (1-412)

Qy 3 ProMetPheTyrPheLeuLeuAlaPheThrThrValLeu---AlaAlaThrAlaAsnAla 21  
Db 46 CCTCTCTTCTTCTTCTTGTGCTTACCTCATATCTACCTTCAGCCAGCGCTGATGAT 105  
Qy 22 GlyProValLeuAspThrAspGlyAspIleIlePheAspGly---SerTyrTrpValLeu 40  
Db 106 GATCATGTGTACGACACTGACGGGCAATAGGTTCAATACGGTGTCAATTATTTCGTGTTG 165  
Qy 41 ProLeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60  
Db 166 CCAGTGATAAGAGGA---AATGGCGGTGGATACAA---GTAGCAAAAGCAGGAACGAA 219  
Qy 61 ---CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79  
Db 220 ACTTGCCTCTGACTGTGGTGCATCTGCAACGAGCTCTCTGAGGGGTTACCGATAAG 279  
Qy 80 PheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeu---AsnIle 98  
Db 280 ATTGCATCTCGATCATCGAGGAGTCTGCTTTCATCACCCAGGCCAGCTTTTCAAGAGCAT 339  
Qy 99 LysMetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAla 118  
Db 340 CAGTTCGGTGTGTTTCCATCCACTCAGACCAGG-----GTGTCCCCAGGT 387  
Qy 119 ProSerPro 121  
Db 388 CCAATTCTCT 396

## RESULT 3

US-09-938-842A-1465  
; Sequence 1465, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRIP1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22



```
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1465
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-630)

Alignment Scores:
Pred. No.: 5,22e-10 Length: 630
Score: 160.50 Matches: 63
Percent Similarity: 43.56% Conservatives: 35
Best Local Similarity: 28.00% Mismatches: 76
Query Match: 14.09% Indels: 51
DB: 9 Gaps: 14

US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-630)
QY 9 LeuAlaPheThrValLeuAlaAlaThrAlaAsnAlaGlyProValLeu----- 25
Db 13 ATATCCATCACTACCACTTTCTCGTGGTGGCTTTGGCGGCACCTCCCTAGCTCGTCT 72
QY 26 -----AspThrAspGlyAspGlyLeuPheAspGly---SerTyrThrVal 39
Db 73 GACAACCATGTCGAGGACTCTGTAGCCGCTCTACTGCTGGTCAACAGTACCACATC 132
QY 40 LeuProLeuIleThrValPro---ThrGlyGlyLeuThrValSerArgArgGly 58
Db 133 GTACCTGCG-----AATCCGACGACGAGGAGGT-----ATTTCGACAGTGA 180
QY 59 AsnGlnCysProLeuPheLeuGlnArgSerGluValAsnArgGlyLeuProVal 78
Db 181 GAAATCTGCTCTTGCATCTTCCAGTCAACAAATCCGCTGACTTGGCTACCCATC 240
QY 79 LysPheSerAsnThrArgSerArgValGlyPheValProGluGluAsnLeuAsn 98
Db 241 AATATT-----AAGTCCGAGTATGTTGTTGAAGAAATGAATGATACCATC 291
QY 99 LysMetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAla 118
Db 292 GAGTTT-----GAGGCTCCGAACTGTTTGTGCTCTAAA 327
QY 119 ProSerPro---TrpArg-----SerLeuPheIleAla 128
Db 328 GAATCCAAGGGTGGAGAGTCTGTACTCTGAAGAATTCAAAGAGTCTTATAATAAGC 387
QY 129 AlaGlyProLysProGluAlaGlyGlyGluAspSerSerArgSerPhePheGlnLeu 148
Db 388 ACT-----GGTGTTCATCAACCCCAAGTGGCTTCAGATCCAT 426
QY 149 LysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSer-----GluGlyAsn 166
Db 427 CGAGTCGACGGA-----GGTCTTACAAGATTGTATATTGTACAAACATCTCGACTACT 480
QY 167 AspCysIleAspValGly---LysAsnGluGluGlyValArgGlyLeuValLeuGly 185
Db 481 ACGTGCATCAAGCTGGCATATATCCAGCATATCTCTGTCGACGACGCTTAGCCCTGACC 540
QY 186 SerThrProProPheAlaThrProPheGluValValPheValLysAlaThrGlyThrAsp 205
Db 541 ACGGATGAGGCTCTCTAGTTAAGTTCAG-----AAGGCA5CAACTCCAAA 588
QY 206 ThrSerSerLysThr 210
Db 589 GCTGATTGTAAGACT 603

RESULT 4
US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-630)
; Sequence 15072, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 15775
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-016-Q1-B1-D11
US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-630)

Alignment Scores:
Pred. No.: 5.28e-10 Length: 396
Score: 158.00 Matches: 48
Percent Similarity: 50.37% Conservatives: 20
Best Local Similarity: 35.56% Mismatches: 59

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 15775
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-016-Q1-B1-D11
US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-630)

Alignment Scores:
Pred. No.: 1.82e-10 Length: 275
Score: 160.00 Matches: 38
Percent Similarity: 59.30% Conservatives: 13
Best Local Similarity: 44.19% Mismatches: 29
Query Match: 14.05% Indels: 6
DB: 10 Gaps: 4

US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-275)
QY 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrVal-----LeuAlaAla 17
Db 12 ATGAAGCTACCTCCCTATATATCCCTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 71
QY 18 ThrAlaAsnAlaGlyProValLeuAspThrAspGlyAspGlyLeuPhe---AspGlySer 36
Db 72 TCAGAAGATGTTGAACAGTTGTGACATAGTGGCAACCCCATTTTCCAGGTGGGACA 131
QY 37 TyrTyrValLeuProLeuIleThrValProGlyGlyLeuThrValSerArg 56
Db 132 TATTACATTTATGCTCACTCAACTTGGGCGCTCCGCTGGTGGATTGACACTA---GGCCGG 188
QY 57 ArgGlyAsn---GlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGly 75
Db 189 ACAGAACTCAAACTGCCAGTACTGTTTTCGAAGATTACTCAGAAATCTTCCGTGGC 248
QY 76 IleProValLysPheSer 81
Db 249 ACACAGTCAAAATTCAGC 266

RESULT 5
US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-275)
; Sequence 3061, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 15775
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-016-Q1-B1-D11
US-09-822-080b-2 (1-214) x US-09-822-080b-2 (1-275)

Alignment Scores:
Pred. No.: 5.28e-10 Length: 396
Score: 158.00 Matches: 48
Percent Similarity: 50.37% Conservatives: 20
Best Local Similarity: 35.56% Mismatches: 59
```

QY	41	ProLeuIleTrpGlyProThrGlyGlyGlyLeuThrLeuValSerArgArgGlyAsnGln	60
		:	
Db	151	CCAGTGATRAGAGGA---AATGCGGTGGAATACAA--GTAGCAAAACAGCAACGAA	204
		:	
QY	61	---CysProLeuPheIleGlyGlnGluArgSerGluuValAsnArgGlyIleProValLys	79
Db	205	ACTTGGCCCTGACTGTGTGTCAACTCTGGCAGCAGCTCTCTGAGGGGTTCGCGATAAAG	264
QY	80	PheSerAsn	82
		:	
Db	265	ATTGCATCT	273

## RESULT 7

```

US-09-878-574-14572
; Sequence 14572, Application US/09878574
; Patent No. US20020110548A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(13401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 14572
; LENGTH: 308
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701068956H1
; US-09-878-574-14572

```

## RESULT 6

```
US-09-878-574-14572
```

Alignment Scores:	
Pred. No.:	7 25e-10
Score:	135.50
Percent Similarity:	69.88%
Best Local Similarity:	44.58%
Query Match:	13.65%
DB:	10
Length:	308
Matches:	37
Conservative:	21
Mismatches:	20
Indels:	5
Gaps:	5

```
US-09-822-080B-2 (1-214) x US-09-878-574-14572 (1-308)
```

Alignment Scores:

```

QY 61 ---CysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79
      ||||| ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 227 ACTTGCCCTCTGACTGTGTGGTCACTCTGGCAACGAGACTCTCTCAGGGGTTACCGATAAAG 286

QY 80 PheSerAsn 82
      :||| :
Db 287 ATTGCATCT 295

RESULT 8
US-09-878-574-14101
; Sequence 14101, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.

```

1

; APPLICANT: Thompson, Michael D.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(15401)B  
 ; CURRENT APPLICATION NUMBER: US/09/878,574  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/333,535  
 ; PRIOR FILING DATE: 1999-06-14  
 ; NUMBER OF SEQ ID NOS: 15775  
 ; SEQ ID NO 14101  
 ; LENGTH: 280  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; OTHER INFORMATION: Clone ID: 701068323H1  
 US-09-878-574-14101

Alignment Scores:  
 Pred. No.: 9.49e-10 Length: 280  
 Score: 154.00  
 Percent Similarity: 58.14% Matches: 37  
 Best Local Similarity: 13.02% Conservative: 13  
 Query Match: 13.52% Mismatches: 30  
 DB: 6 Indels: 6  
 Gaps: 4

US-09-822-080B-2 (1-214) x US-09-878-574-14101 (1-280)

Qy 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrVal-----LeuAlaA17  
 Db 14 ATGAAGCCTACCTTATATATCCCTTCTCTGCTCTCTTTGGTTCCTGGCTCTT 73  
 Qy 18 ThrAlaAsnAlaGlyProValLeuLeuAlaPheThrValSerArg 36  
 Db 74 TCAGAAGATGTTGAACAAGTTGTGACATAAGTGGCAACCCATTTTCCAGGTGGCACA 133  
 Qy 37 TyrTyrValLeuProLeuIleTyrPheLeuLeuAlaPheThrValSerArg 56  
 Db 134 TATTACATTATGTCCTCAACTTGGCGGCTGCGGTGGTGGTGAAGTAACTA---GGCCGG 190  
 Qy 57 ArgGlyAsn---GlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGly 75  
 Db 191 ACAGAAACTCAAACTGCCAGTTACTGTTTCAAGATTACTCAGAAATCTTCCTGGGC 250  
 Qy 76 IleProValLysPheSer 81  
 Db 251 ACACCACTCAAAATTCAGC 268

RESULT 9  
 US-09-878-574-11210  
 ; Sequence 11210, Application US/09878574  
 ; Patent No. US20020110548A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrum, Joseph R.  
 ; APPLICANT: La Rosa, Thomas J.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(15401)B  
 ; CURRENT APPLICATION NUMBER: US/09/878,574  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/333,535  
 ; PRIOR FILING DATE: 1999-06-14  
 ; NUMBER OF SEQ ID NOS: 15775  
 ; SEQ ID NO 11210  
 ; LENGTH: 272  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; OTHER INFORMATION: Clone ID: 701064180H1  
 US-09-878-574-11210

Alignment Scores:  
 Pred. No.: 4.62e-09 Length: 272  
 Score: 148.00 Matches: 36  
 Percent Similarity: 56.98% Conservative: 13

Best Local Similarity: 41.86% Mismatches: 31  
 Query Match: 12.99% Indels: 6  
 DB: 10 Gaps: 4  
 US-09-822-080B-2 (1-214) x US-09-878-574-11210 (1-272)  
 Qy 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrVal-----LeuAlaA17  
 Db 10 ATGAAGCCTACCTTATATATCCCTTCTCTGCTCTCTTTGGTTCCTGGCTCTT 69  
 Qy 18 ThrAlaAsnAlaGlyProValLeuLeuAlaPheThrValSerArg 36  
 Db 70 TCAGAAGATGTCGAACAAGTTGTGACATAAGTGGCAACCCATTTTCCAGGTGGCACA 129  
 Qy 37 TyrTyrValLeuProLeuIleTyrPheLeuLeuAlaPheThrValSerArg 56  
 Db 130 TATTACATTATGTCCTCAACTTGGCGGCTGCGGTGGTGAAGTAACTA---GGCCGG 186  
 Qy 57 ArgGlyAsn---GlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGly 75  
 Db 187 ACAGAAACTCAAACTGCCAGTTACTGTTTCAAGATTACTCAGAAATCTTCCTGGGC 246  
 Qy 76 IleProValLysPheSer 81  
 Db 247 ACACCACTCACATTCAGC 264

RESULT 10  
 US-09-878-574-11929  
 ; Sequence 11929, Application US/09878574  
 ; Patent No. US20020110548A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Byrum, Joseph R.  
 ; APPLICANT: La Rosa, Thomas J.  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; FILE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(15401)B  
 ; CURRENT APPLICATION NUMBER: US/09/878,574  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/333,535  
 ; PRIOR FILING DATE: 1999-06-14  
 ; NUMBER OF SEQ ID NOS: 15775  
 ; SEQ ID NO 11929  
 ; LENGTH: 277  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; OTHER INFORMATION: Clone ID: 701065093H1  
 US-09-878-574-11929

Alignment Scores:  
 Pred. No.: 7.11e-09 Length: 277  
 Score: 146.50 Matches: 33  
 Percent Similarity: 66.67% Conservative: 19  
 Best Local Similarity: 42.31% Mismatches: 23  
 Query Match: 12.86% Indels: 3  
 DB: 10 Gaps: 3  
 US-09-822-080B-2 (1-214) x US-09-878-574-11929 (1-277)  
 Qy 4 MetPheTyrPheLeuLeuAlaPheThrValLeu---AlaAlaThrAlaAsnAlaGly 22  
 Db 42 CTCCTTCTTCTACTTTGTGCTTCCCTCATATCTACTCTCAGCCACCGCTGATGAT 101  
 Qy 23 ProValLeuAspThrAspGlyAspIlePheAspGly---SerTyrTyrValLeuPro 41  
 Db 102 CATGTGTACGACACTGACGGCGATAAGCTTCAATACGGTGTCAATTATTTCGTGTC 161  
 Qy 42 LeuIleTyrPheProThrGlyGlyLeuThrValSerArgGlyAsnGlnCys 61  
 Db 162 GTGATAGAGGA---AATGGCGGTGGAAATCAAGTACAAACGAGGAAACGATATTC 218  
 Qy 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79  
 Db 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIleProValLys 79



Db 257 TATGTTATCGTGACCTTCATTCATT-----GGCACCCAGTTTCATT-----TTA 304  
QY 84 ArgSerArgValcylgPheValProGluGluGluAsnLeuAsnIleIlePheMetAspValGlu 103  
Db 305 GCACCATGCTTGACCATGTTCCACACTCACTGATCAACCATTAATGATTCCTGTG--- 361  
QY 104 ProThrIleCysAlaGlnSerAlaTyrTrp 113  
Db 362 GTGACAGTGTGTAAACAGCCAAACAGTGTGG 391

## RESULT 14

US-09-878-574-11909  
; Sequence 11909, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 11909  
; LENGTH: 282  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701065067H1  
US-09-878-574-11909

Alignment Scores:  
Pred. No.: 5.54e-07 Length: 282  
Score: 130.50 Matches: 34  
Percent Similarity: 59.49% Conservative: 13  
Best Local Similarity: 43.04% Mismatches: 25  
Query Match: 11.46% Indels: 7  
DB: 10 Gaps: 4

US-09-822-080B-2 (1-214) x US-09-878-574-11909 (1-282)

QY 1 MetAsnProMetPheTyr-----PheLeuLeuAlaPheThrThrValLeu----- 15  
Db 47 GTCTCCCTTAGCATTTTCATCTTTTGTCTTCCACATAGAACTCTTCATTGGC 106  
QY 16 AlaAlaThrAlaAsnAlaGlyProValLeuAspThrAspGlyAspIleIlePheAspGly 35  
Db 107 ATTGCTTCAGCAGCAGCAAGAACAGTAGCTGTGACACCTCAGGCCAGAGCTGAGAACAGGT 166  
QY 36 ---SerTyrTyrValLeuProLeuIleTrpGlyProThrGlyGlyGlyLeuThrVal 54  
Db 167 GTCAAGTACTACATTCATACAGTCTTCAGAGCC---AGAGGTGGAGGCCCTAACAGTTTCA 223  
QY 55 SerArgArgGlyAsnGlnCysProLeuPheIleGlyGlnGluArgSerGluValAsn 73  
Db 224 AGCAGTGGCAACAACACATGCCCCCTCTTGTGTGTCAGAGAGAGCTTGAAGTCTCA 280

## RESULT 15

US-09-878-574-69  
; Sequence 69, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 69  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: LIB3028-056-Q1-B1-E6  
US-09-878-574-69

## Alignment Scores:

Pred. No.: 9.87e-07 Length: 384  
Score: 130.00 Matches: 40  
Percent Similarity: 48.80% Conservative: 21  
Best Local Similarity: 32.00% Mismatches: 44  
Query Match: 11.41% Indels: 20  
DB: 10 Gaps: 8

US-09-822-080B-2 (1-214) x US-09-878-574-69 (1-384)

QY 2 AsnProMetPhe-----TyrPheLeuLeuAlaPheThrThrValLeuAlaAlaThr 18  
Db 40 AATACTATCTCTGCTGCTCTTCTTCTTGTGCTTCCACCTCATACTACCTTCA 99  
QY 19 AlaAsnAlaGlyProValLeuAspThrAspGlyAspIleIlePheAsp---GlySerTyr 37  
Db 100 ACCACCGCG-----GTGGTCGATACGAGCGGTGATATCTTCAGAACTCTGCACATAC 153  
QY 38 TyrValLeuProLeuIleTrpGlyProThrGlyGlyGlyLeuThrLeuValSerArgArg 57  
Db 154 TTTATCTTGTCC---GTTTTTCGACCC---GGCGCGGAGTAGAATTCGCCGCCACTGGA 207  
QY 58 GlyAsnGlnCysProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyIlePro 77  
Db 208 AACGAAACTTGCCTCTCACTGTCTGCAG-----ACTCTCTCGGAGGGGCTTTCCA 261  
QY 78 VallysPheSerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsn 97  
Db 262 GCAATATATCGTCC---CGTTACGAATCCCTTTCATCGCGGAGGAGCAACTCTTCAGC 318  
QY 98 IleLysMetAspValGluProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrPro 117  
Db 319 ATCTGTTCGTATTGTGCCA-----TGCTGCGCCACC 351  
QY 118 AlaProSerProTrp 122  
Db 352 ACTCTTCTTAAGTGG 366

Search completed: January 25, 2003, 18:33:53  
Job time : 64 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 25, 2003, 18:33:36 ; Search time 2177 seconds

(without alignments)  
1592.023 Million cell updates/sec

Title: US-09-822-080B-2

Perfect score: 1139

Sequence: 1 MNPFYFLLAFTTILAATAN.....EWFEVKATGDTSSKTSNII 214

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh  
-Q/cgn2\_1/USPTO\_spool/US09822080/runat\_22012003\_142642\_4561/app\_query.fasta\_1.391  
-DB=EST -QFMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09822080.ecgn.1.1.763 @runat\_22012003\_142642\_4561 -NCPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	998	87.6	685	17	BH711815
2	951	83.5	830	17	BH606894
3	479.5	42.1	632	17	BH604620
4	464	40.7	604	10	AV832159
5	456	40.0	660	17	BH456221
6	445	39.1	742	17	BH491802
7	435.5	38.2	726	17	BH457950
8	415	36.4	742	17	BH499547
9	385	33.8	773	17	BH711232
10	360	31.6	374	17	BH514321
11	357.5	31.4	651	17	BH420085
12	352.5	30.9	666	17	BH678808
13	329.5	28.9	566	10	AV545497
14	328.5	28.8	752	17	BH544042
15	327.5	28.5	556	9	AI998304
16	324.5	28.5	566	14	N38333
17	258	22.7	441	14	R30257
18	255.5	22.4	768	10	BE034376
19	252.5	22.2	762	14	BQ165664
20	251.5	22.1	713	12	BG584634
21	246.5	21.6	645	10	AW587317
22	239.5	21.0	796	14	BQ165665
23	237	20.8	648	13	BI421392
24	237	20.8	675	10	AW035922
25	237	20.8	696	13	BI922112
26	237	20.8	731	13	BI921127
27	236.5	20.8	653	10	AW311103
28	236.5	20.8	769	10	BE033927
29	236	20.7	674	10	AW094442
30	232	20.4	695	13	BI423047
31	229	20.1	758	10	AW350211
32	227	19.9	672	13	BM404018
33	226	19.8	648	13	BI432942
34	226	19.8	715	13	BI434637
35	226	19.8	796	14	BQ515969
36	225	19.8	559	10	BE346980
37	224.5	19.7	609	9	AI777200
38	224.5	19.7	629	10	AW030414
39	224.5	19.7	634	10	AW030449
40	224	19.7	563	10	AW567775
41	224	19.7	579	14	BQ297552
42	224	19.7	579	14	BQ453253
43	224	19.7	580	14	BQ454018
44	224	19.7	697	13	BI922845
45	221	19.4	562	10	AW781179

ALIGNMENTS

RESULT 1  
BH711815  
LOCUS BH711815 685 bp DNA linear GSS 20-FEB-2002  
DEFINITION BOHYO04TR BO\_2\_3\_KB Brassica oleracea genomic clone BOHYO04, DNA sequence.  
ACCESSION BH711815 GI:18802293  
VERSION BH711815  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 685)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHY004TF

TIGR Contact: Chris Town

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

source

1. .685

/organism="Brassica oleracea"

/strain="T01000DH3"

/db\_xref="taxon:3712"

/clone="BOHY004"

/clone\_lib="BO\_2\_3\_KB"

/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

176 a 164 c 171 g 174 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 3.74e-106 Length: 685

Score: 998.00 Matches: 186

Percent Similarity: 91.51% Conservative: 8

Best Local Similarity: 87.74% Mismatches: 18

Query Match: 87.62% Indels: 0

DB: 17 Gaps: 0

US-09-822-080B-2 (1-214) x BH711815 (1-685)

Qy 3 PrometPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsnAlagly 22  
Db 1 CCTATGTTTACTTCTTCTTCCCTTAACAGCTGTTTATAGCGCGACGCCAACGACGA 60  
Qy 23 ProValLeuAspThrAspGlyAspLeuLeuPheAspGlySerTyrTyrValLeuProLeu 42  
Db 61 CCAATTTCTGCACATCATGCTGATATATATCCACGCGAGTACTACGTTATCCCGCTC 120  
Qy 43 IleTrrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysPro 62  
Db 121 ATCCGGGGCCCTGAAGTGGCGCTTAACCTCAACCCCGCAACGCAACAGTGTCCC 180  
Qy 63 LeuPheileGlyClnGluArgSerGluValAsnArgGlyLeuProValLysPheSerAsn 82  
Db 181 CTCCTTTATCGGACAGAGCGTTCAGAGGTCGAAAGGGGCAATTCCTCGTGAATTC 240  
Qy 83 TrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMetAspVal 102  
Db 241 TGGAGTCTAGAGTGGGTTCGTTCCGAAATCCGAGAACCTCAACATCAAGATGATATC 300  
Qy 103 GluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrp 122  
Db 301 GAACCTACGTTCTCGCTCAGTCAACTTATTTGGTGGGTCACTACAGCCCCCAGTCCCTG 360  
Qy 123 ArgSerLeuPheileAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerArg 142  
Db 361 AGATCGGGCTTCATAGCGTTGGTCTAAGCCAGAGCTGGAGGAGAAGATTTCGTCGAG 420  
Qy 143 SerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCys 162  
Db 421 AGTTTCTCCAGATCAAGAAACTGAGCAAACTTAACCTTAAAGTTTGTATTCGT 480  
Qy 163 SerGluGlyAsnAspCysIleAspValGlyLysAsnGluClnGlyGlyValArgGlyLeu 182  
Db 481 AGAGACGGTAAACGATTCATGATGCGCTAAACAGAGGAGAGGTGGCGTTCGGGGTTG 540  
Qy 183 ValLeuGlySerThrProPheAlaThrProPheGluValValPheValLysAlaThr 202  
Db 541 GTTTTAGGCTCTAGGCCACCATTCGCTACCCCATTCGAGGGGTGTGTTCTGTGAAGCTACT 600

Qy 203 GlyThrAspThrSerSerLysThrMetSerIlelle 214  
Db 601 GGGTCAGAACTTCATCCAGACTATGCTATATC 636

RESULT 2

BH606894/c

LOCUS

DEFINITION BOGFM83TR BOGF Brassica oleracea genomic clone BOGFM83, DNA

sequence.

ACCESSION BH606894

VERSION BH606894.1 GI:17859340

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 830)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOGFM83TF

COMMENT

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .830

/organism="Brassica oleracea"

/strain="T01000DH3"

/db\_xref="taxon:3712"

/clone="BOGFM83"

/clone\_lib="BOGF"

/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

213 a 185 c 193 g 239 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.53e-100 Length: 830

Score: 951.00 Matches: 179

Percent Similarity: 91.13% Conservative: 6

Best Local Similarity: 88.18% Mismatches: 18

Query Match: 83.49% Indels: 1

DB: 17 Gaps: 0

US-09-822-080B-2 (1-214) x BH606894 (1-830)

Qy 1 MetAsnPrometPheTyrPheLeuLeuAlaPheThrThrValLeuAlaAlaThrAlaAsn 20  
Db 620 ATGAATCCTATGTTTACTTCTTCTTCCCTTAACAGCTGTTTATAGCGCGACGCCAAC 561  
Qy 21 AlaGlyProValLeuAspThrAspGlyAspIlelePheAspGlySerTyrTyrValLeu 40  
Db 560 GCAGGACCAAGTGTTCGACATCGATGGTGGTATCATATATCCACGCGAGTACTACGTTATC 501  
Qy 41 ProLeuIleTrrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGln 60  
Db 500 CCCGTATCCGGGGCCCTGAAGGTGGCGGCTTAACCTCACCACCCCGCAACGCCAACAG 441  
Qy 61 CysProLeuPheileGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80  
Db 440 TGTCCCTCTTTATCGGACAGGCGGTTTCAGAGGTTCGAAAGGGGCAATTCCTCGTGAATTC 381  
Qy 81 SerAsnTrpArgSerArgValGlyPheValProGluGluAsnLeuAsnIleLysMet 100  
Db 380 TCAAACTGGAGGTCTAGAGTTGGGTTGCTTCCCGAATCCGAGAACCTCAACATCAAGATG 321





as reported previously (Seki et al., 1998) cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

# FEATURES

Location/Qualifiers

1..604

/organism="Arabidopsis thaliana"

/db\_xref="taxon:3702"

/clone="RAF11-04-122"

/clone\_lib="RAF111"

/dev\_stage="plants at various developmental stages from germination to mature seeds"

/lab\_host="DH10B"

/note="Site\_1: BamHI; Site\_2: SalI; subjected to various treatments (dehydration, cold, high salt, ABA, heat and UV).

. Dark-grown plants"

153 a 142 c 156 g 152 t 1 others

BASE COUNT

ORIGIN

Alignment Scores:  
Pred. No.: 8.65e-44 Length: 604  
Score: 464.00 Matches: 105  
Percent Similarity: 65.43% Conservative: 18  
Best Local Similarity: 55.85% Mismatches: 55  
Query Match: 40.74% Indels: 10  
DB: 10 Gaps: 6

US-09-822-080B-2 (1-214) x AV832159 (1-604)

QY 1 MetAsnProMetPheTyrPheLeuLeuAlaPheThrValLeuAlaAlaThrAlaAsn 20  
|||||  
DB 50 ATGAATCCTAAGTTTACTTGGTTCCTTAACCGCGGTCTCGCCCTCAACGCA--- 106  
|||||  
QY 21 AlaGlyProValLeuAspThrAspGlyAspIlePheAspGlySerTyrTyrValLeu 40  
|||||  
DB 107 TATGGTCGGTTGACACATCGATCGAAGCCGATGTCACGAAAGTTACTACGTTCTC 166  
|||||  
QY 41 ProteulterpGlyProThrGlyGlyLeuThrLeuValSerArgGlyAsnGln 60  
|||||  
DB 167 CCTGTCATCGGTGGC---CGAGCGGAGGCTGACTCTAGCAGCGCGGTGGGAGCCA 223  
|||||  
QY 61 CysProLeuPheLeGlyGlnGluArgSerGluValAsnArgGlyIleProValLysPhe 80  
|||||  
DB 224 TGTCTTACGATATCGTCAGGAATCTTCAGAAAGTTGATGAGGCGATTCCTCGTAAATTC 283  
|||||  
QY 81 SerAsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnLysMet 100  
|||||  
DB 284 TCAAACTGGAGGCTTAAGTTTGGTTCGTTCCCGAATCAAGAACTCAACATCGAAACA 343  
|||||  
QY 101 AspValGluProThrIleCysAlaGlnSerAlaTyrTrpValThrProAlaProSer 120  
|||||  
DB 344 GACGTCGGAGCCAGATCTCGATCCAGCTCAACCTACTTGGCGGTGGTGGATTGACAC 403  
|||||  
QY 121 ProTrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSer 140  
|||||  
DB 404 GAGAGGAGGAGTACTTCTGGTGGTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463  
|||||  
QY 141 SerArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheVal 160  
|||||  
DB 464 TTGAAGAGTTTCTTCAAGATCGAGAAATCTGGAGAG-----GATGCTTACAAAGTTTGTG 517  
|||||  
QY 161 Phe-----CysSerGluGlyAsn---AspCysIleAspValGly---LysAsn 174  
|||||  
DB 518 TTCGTCTCGGATCTGGCATCTGGCAATCCAAATCCAGGATGTCGGGATATTCATA 577  
|||||  
QY 175 GluGluGlyGlyValArgGlyLeu 182  
|||||  
DB 578 GATGAAGTTCGCGTTCGTCGNTG 601  
|||||  
RESULT 5  
BH456221/c

LOCUS BH456221 660 bp DNA linear GSS 12-DEC-2001  
DEFINITION BOHR52TF BOHR Brassica oleracea genomic clone BOHRQ52, DNA sequence.  
ACCESSION BH456221  
VERSION BH456221.1 GI:17641932  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
REFERENCE 1 (bases 1 to 660)  
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOHRQ52TR  
CONTACT: Chris Town  
TIGR  
7912 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

## FEATURES

Location/Qualifiers

1..660

/organism="Brassica oleracea"

/strain="TOL000DH3"

/db\_xref="taxon:3712"

/clone="BOHRQ52"

/clone\_lib="BOHR"

/note="vector: pHS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHS1 using BstXI linkers"

193 a 163 c 145 g 159 t

BASE COUNT

ORIGIN

Alignment Scores:  
Pred. No.: 8.43e-43 Length: 660  
Score: 456.00 Matches: 106  
Percent Similarity: 62.83% Conservative: 14  
Best Local Similarity: 55.50% Mismatches: 64  
Query Match: 40.04% Indels: 7  
DB: 17 Gaps: 5  
US-09-822-080B-2 (1-214) x BH456221 (1-660)  
QY 26 AspThrAspGlyAspIlePheAspGlySerTyrTyrValLeuProLeuIleTrpGly 45  
|||||  
DB 660 GACATGTGTTGATATCATCAACCAACGCGAGTACTTTGTTCTCCCTCGCCATGCCCGC 601  
|||||  
QY 46 ProThrGlyGlyGlyLeuThrLeuValSerArgArgGlyAsnGlnCysProLeuPheIle 65  
|||||  
DB 600 GCTATTGTTGGCGGCTGAATCTCTCCGCGATCTTGGCTAAATGTCCCATATATT 541  
|||||  
QY 66 GlyGlnGluArgSerGluValAsnArgGlyIleProValLysPheSerAsnTrpArgSer 85  
|||||  
DB 540 GGGCAGGAAGATCCAAACGGTCAACAGGGGCTTCCAGTTAAATTTCAAACTGGCAGTCT 481  
|||||  
QY 86 ArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAspValGluProThr 105  
|||||  
DB 480 AAAGTTGGTTCGTTCCCTGAATCAGAGAGGCTCAACATCGAGATGGATGGCAACCTACG 421  
|||||  
QY 106 IleCysAlaGlnSerAlaTyrTrpValThrProAlaProSerProTrp---ArgSer 124  
|||||  
DB 420 ATTGTGTCACGCCAACCTATTGGTGGGCAT---GGGAGCATCCATGGTTACATTCG 364  
|||||  
QY 125 LeuPheIleAlaAlaGlyProLysProGluAlaGlyGluAspSerSerArgSerPhe 144  
|||||  
DB 363 CTTTCCATAAAGTTGGTCTCTAAGCAGGTGCT---GGAAAGATTCGATCAACCTTTC 307  
|||||  
QY 145 PheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPheCysSerGlu 164  
|||||

```

Db 306 TTTTCAGATCAAGAAACTGAA--GATGTTGGCGTTTACATATATGCGAGTTGCTCTGC 250
Qy 165 GlyAsnAspCysIleAspValGlyLysAsnGluGluGlyValArgGlyLeuValLeu 184
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 249 GGTAAACACTTGGGAGACTGTTGGGTACTTGTGGGTGGTGGCGTTCGACGTTTGGTT--- 193
Qy 185 GlySerThrProPheAlaThrProPheGluValValPheValLysAlaThrGlyThr 204
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 192 ---GCAAGTTTGCATACGCTAAGCCATCCCGGTTCCGGCTCGTGAACCTACTGGGACT 136
Qy 205 Asp-ThrSerLysThrMetSerIleile 214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 TCGTACTTCGTCGAAGACTATGCTATTATC 105

RESULT 6
BH491802
LOCUS BH491802 742 bp DNA linear GSS 13-DEC-2001
DEFINITION BOHIM14TF BOH Brassica oleracea genomic clone BOHIM14, DNA
sequence.
ACCESSION BH491802
VERSION BH491802.1 GI:17699906
SOURCE GSS.
ORGANISM Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 742)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOHIM14TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
    source
        1..742
            /organism="Brassica oleracea"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHIM14"
            /clone_lib="BOH"
            /note="Vector: pHOsl; Site_1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOsl using BstXI linkers"
BASE COUNT 187 a 156 c 179 g 220 t
ORIGIN
Alignment Scores:
Pred. No.: 1,91e-41 Length: 742
Score: 445.00 Matches: 105
Percent Similarity: 63.08% Conservative: 18
Best Local Similarity: 53.85% Mismatches: 54
Query Match: 39.07% Indels: 18
DB: 17 Gaps: 6

US-09-822-080b-2 (1-214) x BH491802 (1-742)

Qy 22 GlyProValLeuAspThrAspGlyAspIleLePheAspGlySerTyrThrValLeuPro 41
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 42 GGTGCTGTTCTTTCACACCGCGGACATCATTTCCGTCGCGAGTTACTATGTTCTCCCC 101
Qy 42 LeuIleTrpGlyProThrGlyGlyLeuThrLeuValSerArgArgGlyAsnGlyCys 61
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 102 GTCTCCGCGGC---CGAGGAGCGCGCTTGTACTAGCGCGCGGTGGGAGCTATGTT 158
Qy 62 ProLeuPheIleGlyGlnGluArgSerGluValAsnArgGlyLeuProValLysPheSer 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 159 CCTTACGACATCGTCGAAGAATCATCTGAACCTGACGACGAAGTATTCCTGTTTCTCG 218
Qy 82 AsnTrpArgSerArgValGlyPheValProGluGluGluAsnLeuAsnIleLysMetAsp 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 AACTGGAGACTAGAGTTGGGTTTGTCTGAGTCACAGACCTTAACATCAAGACGGAC 278
Qy 102 ValClnProThrIleCysAlaGlnSerAlaTyrTrpTrpValThrProAlaProSerPro 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 GTTGAAGCTACGATATGCTTCCAGTCAACATCTGGAGTGGAGTGGTGTGACGAGGAG 338
Qy 122 TrpArgSerLeuPheIleAlaAlaGlyProLysProGluAlaGlyGlyGluAspSerSer 141
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 339 AGCAGCAGTATTTCGTGGTGGCTGCTACAA-----GACGACTCACCC 383
Qy 142 ArgSerPheGlnIleLysLysThrGluAlaLysLeuAsnAlaTyrLysPheValPhe 161
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 AACAGTTCTTCTCCAGATCGAAAATCTGGAGAT-----GATGCTTACAAAGTTTGTTC 437
Qy 162 Cys-----SerGluGlyAsnAspCysIleAspValGly---LysAsnGluGlu 176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 TGTCTCTGCTACTGCTGATTCCTGCTCAATGCGAGGACGTTGGAATATTTGTGGAGGAA 497
Qy 177 GlyGlyValArgGlyLeuValLeuGlySerThrProPheAlaThrProPheGluVal 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 498 ATAGCGCTTCGGGTTTGGCTTTAAGGCTC-----GAGCGGTTCTTGGTT 542
Qy 197 ValPheValLysAlaThrGlyThrAspThrSerSerLysThrMet 211
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db 543 ATGTTCAAGAAAGCTAATGTTACCGAGATATCATCCGAGACTATG 587

RESULT 7
BH457950/c
LOCUS BH457950 726 bp DNA linear GSS 12-DEC-2001
DEFINITION BOHAR95TR BOHA Brassica oleracea genomic clone BOHAR95, DNA
sequence.
ACCESSION BH457950
VERSION BH457950.1 GI:17643661
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 726)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: BOHAR95TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES
    source
        1..726
            /organism="Brassica oleracea"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHAR95"
            /note="Vector: pHOsl; Site_1: BstXI; 2-3 kb sheared
            genomic DNA inserted into pHOsl using BstXI linkers"
BASE COUNT 207 a 145 c 160 g 214 t
ORIGIN
Alignment Scores:
Pred. No.: 2.4e-40 Length: 726
Score: 435.50 Matches: 91
Percent Similarity: 70.00% Conservative: 14

```







[illegible]

## RESULT 15

AI998304/c

## LOCUS

DEFINITION 556 bp mRNA linear EST 08-SEP-1999  
thaliana cDNA clone 701545163, rosette-2 Arabidopsis

## ACCESSION

AI998304

## VERSION

AI998304.1 GI:5845209

## KEYWORDS

EST.

## SOURCE

thale cress.

## ORGANISM

Arabidopsis thaliana

## REFERENCE

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 556)

Chen, J., Momiya, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,

Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,

Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.

, Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,

Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murry, L.,

Turner, C., Krikorian, S., Elder, L. and Hanson, D.

Arabidopsis thaliana Gene Expression MicroArray

Unpublished (1999)

Contact: David Smoller, Ph.D.

Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.

4633 World Parkway Circle, St. Louis, MO 63134, USA

Tel: 877-577-2733

Fax: 314-427-3324

Email: service@genomesystems.com.

Location/Qualifiers

1. 556

/organism="Arabidopsis thaliana"

/cultivar="Columbia Col-0"

/db\_xref="taxon:3702"

/clone\_lib="A. thaliana, Columbia Col-0, rosette-2"

/tissue\_type="rosette"

/dev\_stage="4 - 7 weeks"

/note="Vector: pSPORT; Site\_1: NotI; Site\_2: SalI; cDNA

library was derived from untreated rosette tissue from

Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.

Plants were grown in 1:1:1 peat moss/vermiculite/perlite

soil at 22 deg. C +/- 3 deg. C under constant light, and

watered with fertilizer. cDNA synthesis was initiated

using a NotI-oligo(dT) primer. Double-stranded cDNA was

blunted, ligated to SalI adaptors, digested with NotI,

size-selected, and cloned into the NotI and SalI sites of

the pSPORT vector."

BASE COUNT 159 a 128 c 113 g 155 t 1 others

## ORIGIN

## Alignment Scores:

Pred. No.: 7,05e-28 Length: 556

Score: 327.50 Matches: 76

Percent Similarity: 62.33% Conservative: 15

Best Local Similarity: 52.05% Mismatches: 42

Query Match: 28.75% Indels: 13

DB: 9 Gaps: 5

US-09-822-080B-2 (1-214) x AI998304 (1-556)

QY 72 ValAsnArgGlyIleProValLysPheSerAsnTrpArgSerArgValGlyPheValpro 91

||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 554 GTTGATGAGGGGCAATCCCGTAAATCTCAAACTGGAGGCTTAAGGTTGCGTTCCTCC 495

QY 92 GluGluGluAsnLeuAsnIleLysMetAspValGluProThrIleCysAlaGlnSerAla 111

||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 494 GAATCACAGAACTCAACATCGAAACACAGCGCGAGCCAGCATCTCCATCCCAACC 435

QY 112 TyrTrpTrpValThrProAlaProSerProTrpArgSerLeuPheIleAlaGlyPro 131

||||| ||| :||| :||| :||| :||| :||| :||| :||| :|||

Db 434 TACTGGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTCCA 375

QY 132 LysProGluAlaGlyGlyGluAspSerSerArgSerPheGlnIleLysLysThrGlu 151  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 374 AAGCCAGAAGGGTTCGGACAAGATTTCGTTGAAGATTTCCTCAAGATCGAGAAATCTGGA 315

QY 152 AlaLysLeuAsnAlaTyrIlyPheValPhe-----CysSerGluGlyAsn--- 166  
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 314 GAG-----GATGCTTACAAGTTTGTCTCTCGGACTTCGGACTCTGGCAATCCA 261

QY 167 AspCysIleAspValGly---LysAsnGluGluGlyGlyValArgGlyLeuValLeuGly 185  
||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 260 AAATGCACCGATGTCGGATATTCATAGTGAACCTGGCGTTCGCTTTGGCTTTAAGC 201

QY 186 SerThrProProPheAlaThrProPheGluValValPheValLysAlaThrGlyThrAsp 205  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 200 GAT-----AAGCCGCTCTCTGGTTATGTTCAAAAAGCTAATGTGACCGAA 156

QY 206 ThrSerSerLysThrMet 211  
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 155 GTTTCGTCGAAGACTATG 138

Search completed: January 25, 2003, 21:00:10  
Job time : 2190 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 15:49:28 ; Search time 2850 Seconds  
(without alignments)  
8261.112 Million cell updates/sec

Title: US-09-822-080B-1  
Perfect score: 809  
Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.pa.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rpd.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em\_sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	783	96.8	783	8	BOU18995	U18995 Brassica ol
2	36	4.4	3820	9	IR2004442	AL389982 Homo sapi
3	36	4.4	61920	2	AC021270	AC021270 Homo sapi
4	36	4.4	139623	2	AC040932	AC040932 Homo sapi
5	36	4.4	183443	2	AC068145	AC068145 Homo sapi
6	35	4.3	1084	5	OMY250180	AJ250180 Oncorhync
7	35	4.3	61373	2	AC073453	AC073453 Homo sapi
8	35	4.3	72632	2	AC021344	AC021344 Homo sapi
9	35	4.3	160197	2	AC009814	AC009814 Homo sapi
10	35	4.3	162295	2	AC013768	AC013768 Homo sapi
11	34	4.2	723	9	AF361887S5	AF361891 Homo sapi
12	34	4.2	1263	8	AF230276	AF230276 Triphysar
13	34	4.2	4609	9	HSM803416	AL832109 Homo sapi
14	34	4.2	80105	9	AC004927	AC004927 Homo sapi
15	34	4.2	100680	9	AC011461	AC011461 Homo sapi
16	34	4.2	164650	2	AC093016	AC093016 Homo sapi
17	34	4.2	169546	2	AC004157	AC004157 Plasmodiu
18	34	4.2	201508	2	AC026290	AC026290 Homo sapi
19	34	4.2	323263	2	AC079406	AC079406 Homo sapi
20	33	4.1	274	8	AB030142	AB030142 Pinus den
21	33	4.1	1087	9	BC017948	BC017948 Homo sapi
22	33	4.1	3263	10	BC005471	BC005471 Mus muscu
23	33	4.1	40885	2	AC100188	AC100188 Mus muscu
24	33	4.1	49499	9	AF184110	AF184110 Homo sapi
25	33	4.1	51288	9	AL627210	AL627210 Human DNA
26	33	4.1	75540	2	AC026263	AC026263 Homo sapi
27	33	4.1	94422	9	HSJ575121	AL096841 Human DNA
28	33	4.1	101720	9	AC005006	AC005006 Homo sapi
29	33	4.1	126837	9	AL353652	AL353652 Human DNA
30	33	4.1	131359	9	AC004883	AC004883 Homo sapi
31	33	4.1	146473	2	AC013344	AC013344 Homo sapi
32	33	4.1	149563	9	AC116348	AC116348 Homo sapi
33	33	4.1	150216	9	AC092923	AC092923 Homo sapi
34	33	4.1	150548	2	AC069230	AC069230 Homo sapi
35	33	4.1	153803	2	AC009617	AC009617 Homo sapi
36	33	4.1	154536	2	AC116813	AC116813 Mus muscu
37	33	4.1	154917	9	AC090842	AC090842 Homo sapi
38	33	4.1	156348	9	AC099513	AC099513 Homo sapi
39	33	4.1	158420	9	AL732326	AL732326 Human DNA
40	33	4.1	158513	2	AC011054	AC011054 Homo sapi
41	33	4.1	162209	9	CNS01RTH	AL163195 Human chr
42	33	4.1	162337	2	AC025354	AC025354 Homo sapi
43	33	4.1	164302	9	AC127457	AC127457 Homo sapi
44	33	4.1	165970	10	AL672095	AL672095 Mouse DNA
45	33	4.1	169470	2	AC019291	AC019291 Homo sapi

ALIGNMENTS

RESULT 1  
BOU18995  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

BOU18995 783 bp mRNA linear PLN 04-SPP-1997  
Brassica oleracea trypsin inhibitor propeptide mRNA, complete cds.  
U18995  
U18995.1 GI:841207

Brassica oleracea.  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 783)  
Williams,D.L., Kain,W.C. and Broadway,R.M.  
Isolation and characterization of a serine proteinase inhibitor

```
JOURNAL cDNA (Accession No. U18995) from cabbage (PGR97-083)
REFERENCE 2 (bases 1 to 783)
AUTHORS Williams,D.L.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1994) David L. Williams, Entomology, NYSAES
Cornell University, Geneva, NY 14456, USA
FEATURES
    source
        1..783
            /organism="Brassica oleracea"
            /strain="Superpack"
            /db_xref="taxon:3712"
            /clone_lib="lambda Zap II"
            /dev_stage="mature cabbage"
            2..646
                /citation=[1]
                /codon_start=1
                /product="trypsin inhibitor peptide"
                /protein_id="AAB68964.1"
                /db_xref="GI:841208"
                /translation="MNPFYFLAFTTVAATANAGPVLDTDGDIIFDGSGYVVLPLIW
                GPTGGGLVSRNQCPLFIQERSEVNRGIPVFSNWRVGFQFPEENLNKMDV
                EPTCAQAYWVITPAPSPWRSLFIAAGPFEAGDESSRSFFQFKKTEAKLNAYKFV
                FCSGNDICIDVGKNEBEGVRLVGLSTPPFATPEFVFKATGDTSSKTMSTII"
    sig_peptide
        194 a 186 c 181 g 222 t
    BASE COUNT
        194 a 186 c 181 g 222 t
    CDS
        1..783
            /product="trypsin inhibitor peptide"
            /protein_id="AAB68964.1"
            /db_xref="GI:841208"
            /translation="MNPFYFLAFTTVAATANAGPVLDTDGDIIFDGSGYVVLPLIW
            GPTGGGLVSRNQCPLFIQERSEVNRGIPVFSNWRVGFQFPEENLNKMDV
            EPTCAQAYWVITPAPSPWRSLFIAAGPFEAGDESSRSFFQFKKTEAKLNAYKFV
            FCSGNDICIDVGKNEBEGVRLVGLSTPPFATPEFVFKATGDTSSKTMSTII"
    Query Match 96.8%; Score 783; DB 8; Length 783;
    Best Local Similarity 100.0%; Pred. No. 0;
    Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY 1 GATGAATCTATGTTTACTCTCTTCTGCTTACCACTGTTTGGCGGCGACGCAAA 60
    DB 1 GATGAATCTATGTTTACTCTCTTCTGCTTACCACTGTTTGGCGGCGACGCAAA 60
    QY 61 CGCTGACCACTTCGACACTGATGGTGAATATATTCGACGCGAGTTACTAGTTCT 120
    DB 61 CGCTGACCACTTCGACACTGATGGTGAATATATTCGACGCGAGTTACTAGTTCT 120
    QY 121 CCCCTCATCTGGGCGCTTACAGGTGGCGGCTAACTCTCTCTCCGCTGCGGCAACA 180
    DB 121 CCCCTCATCTGGGCGCTTACAGGTGGCGGCTAACTCTCTCTCCGCTGCGGCAACA 180
    QY 181 GTGTCCTCTCTTATCGGACAGAGCGTTCAGAGTCAACAGGGGCATTCCTCGTGAAT 240
    DB 181 GTGTCCTCTCTTATCGGACAGAGCGTTCAGAGTCAACAGGGGCATTCCTCGTGAAT 240
    QY 241 CTCAACTCGAGGTCAGAGTTGGTTCGTCCTCCCGAAGAAGAACCTCAACATCAAGAT 300
    DB 241 CTCAACTCGAGGTCAGAGTTGGTTCGTCCTCCCGAAGAAGAACCTCAACATCAAGAT 300
    QY 301 GATGTCGAACCTAGATCTGCGCTCAGTCAGTTATTTGGTGGGTCAGTCAGCCCCCAG 360
    DB 301 GATGTCGAACCTAGATCTGCGCTCAGTCAGTTATTTGGTGGGTCAGTCAGCCCCCAG 360
    QY 361 TCCTCGAGGTCGTTGTTTCATAGCGGCTGCTTAAGCCAGAGAGTTCGAGGAGAGATTC 420
    DB 361 TCCTCGAGGTCGTTGTTTCATAGCGGCTGCTTAAGCCAGAGAGTTCGAGGAGAGATTC 420
    QY 421 GTCGAGGAGTTCTTCCAGATCAAGAAACTGAAGCCAACTTAACGCTTACAAGTTTGT 480
    DB 421 GTCGAGGAGTTCTTCCAGATCAAGAAACTGAAGCCAACTTAACGCTTACAAGTTTGT 480
    QY 481 ATCTGTAGTGGGTGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
    DB 481 ATCTGTAGTGGGTGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
    QY 541 GGGTTGGTTTGGTCTACGCCACCACTTCGCTACCCCATTCGAGGTTGTTGCTGCTGAA 600
    DB 541 GGGTTGGTTTGGTCTACGCCACCACTTCGCTACCCCATTCGAGGTTGTTGCTGCTGAA 600
    QY 601 AGCTACTGGGACACACTTTCATCCAGACTATGCTATATCTGAGAGAAATTAAGAC 660
```

```
Db 601 AGCTACTGGGACACACTTTCATCCAGACTATGCTATATCTGAGAGAAATTAAGAC 660
QY 661 CACTTAATAAAGAGGATAAGTGTATAACTTACCTCTCTAATAATAAACTCTATCTATGA 720
Db 661 CACTTAATAAAGAGGATAAGTGTATAACTTACCTCTCTAATAATAAACTCTATCTATGA 720
QY 721 TGATGTTTTCTTTTTCATGATCATCATCATCATCATCATCATCATCATCATCATCATCAT 780
Db 721 TGATGTTTTCTTTTTCATGATCATCATCATCATCATCATCATCATCATCATCATCATCAT 780
QY 781 TCT 783
Db 781 TCT 783

RESULT 2
IR2004442
LOCUS Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 200442.
DEFINITION AL389982
ACCESSION AL389982.1 GI:9367831
VERSION FLCDNA.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3820)
Auffray,C., Ansoorge,W., Ballabio,A., Estivill,X., Gibson,K.,
Lehrach,H., Poustka,A. and Lundeberg,J.
The European IMAGE consortium for integrated Molecular analysis of
human gene transcripts
Unpublished
2 (bases 1 to 3820)
Pluvinet,R., Estivill,X., Escarceller,M. and Sumoy,L.
Direct Submission
Submitted (16-JUL-2000) Dept. Genetica Molecular, Institut de
Recerca Oncologica (IRO), Hospital Duran i Reynals, Av. Gran Via
8/n Km 2,7 L'Hospitalet de Llobregat, 08907 Barcelona, Catalunya,
SPAIN. Tel: ++34-93-260-7775 Fax: ++34-93-260-7776 WWW site:
http://www.iro.es e-mail enquiries: lsumoy@iro.es
EURO-IMAGE Consortium Contact: Auffray C
CNRS UPR 420 - Genetique Moleculaire et Biologie du Developpement
IPR 1221 - Rue Guy Moquet 19, Batiment G - BP 8
94801 Villejuif Cedex, FRANCE
Tel: ++33-1-49 58 34 98
Fax: ++33-1-49 58 35 09
e-mail: auffray@infobiogen.fr
This clone is available royalty-free through IMAGE Consortium
distributors.
IMPORTANT: This sequence represents the full insert of this IMAGE
cDNA clone. No attempt has been made to verify whether this
corresponds to the full-length of the original mRNA from which it
was derived.
FEATURES
    source
        1..3820
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="IMAGE cDNA clone 200442"
            /clone_lib="NCI_CGAP_Pan1"
            /note="appears to have a retained intron from nt positions
            1 to 1446-does not match original EST Acc. No. AI247293"
            <1447..1498
                /codon_start=2
                /product="AK001912 hypothetical protein"
                /protein_id="CAB97537.1"
                /db_xref="GI:9367832"
                /translation="TGQPSAPGDTSVNGPV"
    CDS
        1063 a 753 c 806 g 1198 t
    BASE COUNT
        1063 a 753 c 806 g 1198 t
    Query Match 4.4%; Score 36; DB 9; Length 3820;
```

Best Local Similarity 100.0%; Pred. No. 8.8e-09;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 CTTTGTCTTCTAAAAA...AAAAAAAAA 809  
Db 3701 CTTTGTCTTCTAAAAA...AAAAAAAAA 3736

## RESULT 3

AC021270/c

LOCUS

AC021270 Homo sapiens clone RP11-15D16, LOW-PASS SEQUENCE SAMPLING.  
AC021270

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC021270 61920 bp DNA linear HTG 13-JUL-2000  
Homo sapiens clone RP11-15D16, LOW-PASS SEQUENCE SAMPLING.  
AC021270  
HTG: HTGS\_PHASE0.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 61920)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome, clone RP11-15D16  
Unpublished  
2 (bases 1 to 61920)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fencost, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Lechoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,  
Zimmer, A. and Zody, M.

## Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 3, 2000 this sequence version replaced 61:6705842.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L3444

Center clone name: 15\_D\_16

-----

\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 788: contig of 788 bp in length  
\* 789 888: gap of 100 bp  
\* 889 1659: contig of 771 bp in length  
\* 1660 1759: gap of 100 bp  
\* 1760 2519: contig of 760 bp in length

\* 2520 2619: gap of 100 bp  
\* 2620 3386: contig of 777 bp in length  
\* 3397 3496: gap of 100 bp  
\* 3497 4274: contig of 778 bp in length  
\* 4275 4374: gap of 100 bp  
\* 4375 5162: contig of 788 bp in length  
\* 5163 5262: gap of 100 bp  
\* 5263 6039: contig of 777 bp in length  
\* 6040 6139: gap of 100 bp  
\* 6140 6929: contig of 790 bp in length  
\* 6930 7029: gap of 100 bp  
\* 7030 7824: contig of 795 bp in length  
\* 7825 7924: gap of 100 bp  
\* 7925 8700: contig of 776 bp in length  
\* 8701 8800: gap of 100 bp  
\* 8801 9594: contig of 794 bp in length  
\* 9595 9694: gap of 100 bp  
\* 9695 10483: contig of 789 bp in length  
\* 10484 10583: gap of 100 bp  
\* 10584 11359: contig of 776 bp in length  
\* 11360 11459: gap of 100 bp  
\* 11460 12228: contig of 769 bp in length  
\* 12229 12328: gap of 100 bp  
\* 12329 13095: contig of 767 bp in length  
\* 13096 13195: gap of 100 bp  
\* 13196 13992: contig of 797 bp in length  
\* 13993 14092: gap of 100 bp  
\* 14093 14867: contig of 775 bp in length  
\* 14868 14967: gap of 100 bp  
\* 14968 15756: contig of 789 bp in length  
\* 15757 15856: gap of 100 bp  
\* 15857 16650: contig of 794 bp in length  
\* 16651 16750: gap of 100 bp  
\* 16751 17546: contig of 796 bp in length  
\* 17547 17646: gap of 100 bp  
\* 17647 18437: contig of 791 bp in length  
\* 18438 18537: gap of 100 bp  
\* 18538 19341: contig of 804 bp in length  
\* 19342 19441: gap of 100 bp  
\* 19442 20229: contig of 788 bp in length  
\* 20230 20329: gap of 100 bp  
\* 20330 21120: contig of 791 bp in length  
\* 21121 21220: gap of 100 bp  
\* 21221 22017: contig of 797 bp in length  
\* 22018 22117: gap of 100 bp  
\* 22118 22899: contig of 782 bp in length  
\* 22900 22999: gap of 100 bp  
\* 22999 23795: contig of 795 bp in length  
\* 23795 23894: gap of 100 bp  
\* 23895 24698: contig of 804 bp in length  
\* 24699 24798: gap of 100 bp  
\* 24799 25581: contig of 783 bp in length  
\* 25582 25681: gap of 100 bp  
\* 25682 26465: contig of 784 bp in length  
\* 26466 26565: gap of 100 bp  
\* 26566 27341: contig of 776 bp in length  
\* 27342 27441: gap of 100 bp  
\* 27442 28235: contig of 794 bp in length  
\* 28236 28335: gap of 100 bp  
\* 28336 29115: contig of 780 bp in length  
\* 29116 29215: gap of 100 bp  
\* 29216 29996: contig of 781 bp in length  
\* 29997 30096: gap of 100 bp  
\* 30097 30874: contig of 778 bp in length  
\* 30875 30974: gap of 100 bp  
\* 30975 31763: contig of 789 bp in length  
\* 31764 31863: gap of 100 bp  
\* 31864 32640: contig of 777 bp in length  
\* 32641 32740: gap of 100 bp  
\* 32741 33528: contig of 788 bp in length  
\* 33529 33628: gap of 100 bp  
\* 33629 34414: contig of 786 bp in length  
\* 34415 34514: gap of 100 bp



```

* be preserved.
* 1 1391: contig of 1391 bp in length
* 1392 1491: gap of 100 bp
* 1492 2894: contig of 1403 bp in length
* 2895 2994: gap of 100 bp
* 2995 4882: contig of 1888 bp in length
* 4883 4982: gap of 100 bp
* 4983 7536: contig of 2554 bp in length
* 7537 7636: gap of 100 bp
* 7637 8996: contig of 1360 bp in length
* 8997 9096: gap of 100 bp
* 9097 10646: contig of 1550 bp in length
* 10647 10746: gap of 100 bp
* 10747 13776: contig of 3030 bp in length
* 13777 13876: gap of 100 bp
* 13877 16339: contig of 2463 bp in length
* 16340 16439: gap of 100 bp
* 16440 19281: contig of 2842 bp in length
* 19282 19381: gap of 100 bp
* 19382 23168: contig of 3787 bp in length
* 23169 23268: gap of 100 bp
* 23269 27628: contig of 4360 bp in length
* 27629 27728: gap of 100 bp
* 27729 33725: contig of 5997 bp in length
* 33726 33825: gap of 100 bp
* 33826 39190: contig of 5365 bp in length
* 39191 39290: gap of 100 bp
* 39291 46458: contig of 7168 bp in length
* 46459 46558: gap of 100 bp
* 46559 53598: contig of 7040 bp in length
* 53599 53698: gap of 100 bp
* 53699 62449: contig of 8751 bp in length
* 62450 62549: gap of 100 bp
* 62550 70473: contig of 7924 bp in length
* 70474 70573: gap of 100 bp
* 70574 81067: contig of 10494 bp in length
* 81068 81167: gap of 100 bp
* 81168 93692: contig of 12525 bp in length
* 93693 93792: gap of 100 bp
* 93793 106062: contig of 12270 bp in length
* 106063 106162: gap of 100 bp
* 106163 121295: contig of 15133 bp in length
* 121296 121395: gap of 100 bp
* 121396 139623: contig of 18228 bp in length.
Location/Qualifiers
1. 139623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTD-2014N11"
/clone_lib="CITD Human BAC"
1. 1391
/note="assembly_fragment"
1492. 2894
/note="assembly_fragment"
2995. 4882
/note="assembly_fragment"
4983. 7536
/note="assembly_fragment"
7637. 8996
/note="assembly_fragment"
9097. 10646
/note="assembly_fragment"
10747. 13776
/note="assembly_fragment"
13877. 16339
/note="assembly_fragment"
16440. 19281
/note="assembly_fragment"
19382. 23168
/note="assembly_fragment"
23269. 27628

```

```

misc_feature
27729..33725
/note="assembly_fragment"
misc_feature
33826..39190
/note="assembly_fragment"
misc_feature
39291..46458
/note="assembly_fragment"
misc_feature
46559..53598
/note="assembly_fragment"
misc_feature
53699..62449
/note="assembly_fragment"
misc_feature
62550..70473
/note="assembly_fragment"
misc_feature
70574..81067
/note="assembly_fragment"
misc_feature
81168..93692
/note="assembly_fragment"
misc_feature
93793..106062
/note="assembly_fragment"
misc_feature
106163..121295
/note="assembly_fragment"
misc_feature
121396..139623
/note="assembly_fragment"
BASE COUNT 40549 a 28119 c 27659 g 41192 t 2104 others
ORIGIN
Query Match 4.4%; Score 36; DB 2: Length 139623;
Best Local Similarity 100.0%; Pred. No. 7.9e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 774 CTTGTTTCTAAAAA.....AAAAAAAAAAAAAAAAAAAA 809
|||||.....
Db 121606 CTTGTTTCTAAAAA.....AAAAAAAAAAAAAAAAAAAA 121571

```

# RESULT 5 AC068145/c

```

LOCUS      AC068145      183443 bp      DNA      linear      HTG 04-MAY-2001
DEFINITION Homo sapiens chromosome 17 clone CTD-2531H7 map 17, WORKING DRAFT
SEQUENCE   AC068145      48 unordered pieces.
AC068145    AC068145.3 GI:99666946
VERSION     HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS    Homo sapiens.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 183443)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 17, clone CTD-2531H7
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 183443)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavskiy,L., Bouckghalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Levine,R., LaRocque,K., Lamazares,R., Landers,T., Lehoczeky,J.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrum,J., Meneur,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivier,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.

```

## FEATURES source

```

misc_feature
1492..2894
/note="assembly_fragment"
misc_feature
2995..4882
/note="assembly_fragment"
misc_feature
4983..7536
/note="assembly_fragment"
misc_feature
7637..8996
/note="assembly_fragment"
misc_feature
9097..10646
/note="assembly_fragment"
misc_feature
10747..13776
/note="assembly_fragment"
misc_feature
13877..16339
/note="assembly_fragment"
misc_feature
16440..19281
/note="assembly_fragment"
misc_feature
19382..23168
/note="assembly_fragment"
misc_feature
23269..27628

```

TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (29-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 5, 2000 this sequence version replaced gi:7960325.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L10026  
Center clone name: 2531.H.7  
----- Summary Statistics  
Sequencing vector: M13; W77815; 99% of reads  
Sequencing vector: Plasmid; n/a; 0.0% of reads  
0.758725341426404Chemistry: Dye-terminator Big Dye; 100% of  
reads

Assembly program: Phrap; version 0.960731  
Consensus quality: 152281 bases at least Q40  
Consensus quality: 167680 bases at least Q30  
Consensus quality: 174098 bases at least Q20  
Insert size: 200000; agarose-fp  
Insert size: 178743; sum-of-contigs  
Quality coverage: 2.9 in Q20 bases; agarose-fp  
Quality:  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 48 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1059: contig of 1059 bp in length  
1060 1159: gap of 100 bp  
1160 2777: contig of 1618 bp in length  
2778 2877: gap of 100 bp  
2878 4283: contig of 1406 bp in length  
4284 4383: gap of 100 bp  
4384 5447: contig of 1064 bp in length  
5448 5547: gap of 100 bp  
5548 6776: contig of 1229 bp in length  
6777 6876: gap of 100 bp  
6877 8215: contig of 1339 bp in length  
8216 8315: gap of 100 bp  
8316 9961: contig of 1646 bp in length  
9962 10061: gap of 100 bp  
10062 12224: contig of 2163 bp in length  
12225 12324: gap of 100 bp  
12325 13794: contig of 1470 bp in length  
13795 13894: gap of 100 bp  
13895 16857: contig of 2963 bp in length  
16858 16957: gap of 100 bp  
16958 18266: contig of 1309 bp in length  
18267 18366: gap of 100 bp  
18367 20190: contig of 1824 bp in length  
20191 20290: gap of 100 bp  
20291 21460: contig of 1170 bp in length  
21461 21560: gap of 100 bp  
21561 23407: contig of 1847 bp in length  
23408 23507: gap of 100 bp  
23508 25455: contig of 1948 bp in length  
25456 25555: gap of 100 bp  
25556 27331: contig of 1776 bp in length  
27332 27431: gap of 100 bp  
27432 30406: contig of 2975 bp in length  
30407 30506: gap of 100 bp  
30507 32805: contig of 2299 bp in length  
32806 32905: gap of 100 bp  
32906 35659: contig of 2754 bp in length

35660 35759: gap of 100 bp  
35760 37974: contig of 2215 bp in length  
37975 38074: gap of 100 bp  
38075 40531: contig of 2447 bp in length  
40522 40621: gap of 100 bp  
40622 43219: contig of 2598 bp in length  
43220 43319: gap of 100 bp  
43320 46161: contig of 2842 bp in length  
46162 46261: gap of 100 bp  
46262 48463: contig of 2202 bp in length  
48464 48563: gap of 100 bp  
48564 51275: contig of 2712 bp in length  
51276 51375: gap of 100 bp  
51376 53522: contig of 2147 bp in length  
53523 53622: gap of 100 bp  
53623 56955: contig of 3333 bp in length  
56956 57055: gap of 100 bp  
57056 59555: contig of 2500 bp in length  
59556 59655: gap of 100 bp  
59656 62713: contig of 3058 bp in length  
62714 62813: gap of 100 bp  
62814 66519: contig of 3706 bp in length  
66520 66619: gap of 100 bp  
66620 68930: contig of 2311 bp in length  
68931 69030: gap of 100 bp  
69031 72724: contig of 3694 bp in length  
72725 72824: gap of 100 bp  
72825 76092: contig of 3268 bp in length  
76093 76192: gap of 100 bp  
76193 80143: contig of 3951 bp in length  
80144 80243: gap of 100 bp  
80244 84957: contig of 4714 bp in length  
84958 85057: gap of 100 bp  
85058 88424: contig of 3367 bp in length  
88425 88524: gap of 100 bp  
88525 93428: contig of 4904 bp in length  
93429 93528: gap of 100 bp  
93529 97177: contig of 3649 bp in length  
97178 97277: gap of 100 bp  
97278 102713: contig of 5436 bp in length  
102714 102813: gap of 100 bp  
102814 109141: contig of 6328 bp in length  
109142 109241: gap of 100 bp  
109242 114513: contig of 5272 bp in length  
114514 114613: gap of 100 bp  
114614 124525: contig of 9912 bp in length  
124526 124625: gap of 100 bp  
124626 132364: contig of 7739 bp in length  
132365 132464: gap of 100 bp  
132465 139488: contig of 7024 bp in length  
139489 139588: gap of 100 bp  
139589 148734: contig of 9146 bp in length  
148735 148834: gap of 100 bp  
148835 159465: contig of 10631 bp in length  
159466 159565: gap of 100 bp  
159566 172685: contig of 13120 bp in length  
172686 172785: gap of 100 bp  
172786 183443: contig of 10658 bp in length.

## FEATURES

## source

1. .183443  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="CTD-2531H7"  
/clone\_lib="CITD Human BAC"  
1. .1059  
/note="assembly\_fragment"  
1160. .2777  
/note="assembly\_fragment"  
2878. .4283  
/note="assembly\_fragment"  
4384. .5447

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

```

/note="assembly_fragment"
5548..6776
/note="assembly_fragment"
6877..8215
/note="assembly_fragment"
8316..9961
/note="assembly_fragment"
10062..12224
/note="assembly_fragment"
12325..13794
/note="assembly_fragment"
13895..16857
/note="assembly_fragment"
16958..18266
/note="assembly_fragment"
18367..20190
/note="assembly_fragment"
20291..21460

Query Match      4.4%; Score 36; DB 2; Length 183443;
Best Local Similarity 100.0%; Pred. No. 7.8e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 774 CTTGTGTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
|||||...|||||
Db 108842 CTTGTGTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 108807

RESULT 6
OMY250180      1084 bp      mRNA      linear      VRT 09-OCT-2000
LOCUS          Oncorhynchus mykiss mRNA for beta thymosin.
DEFINITION     AJ250180
ACCESSION      AJ250180.1 GI:10798667
VERSION        beta thymosin.
KEYWORDS       rainbow trout.
SOURCE         Oncorhynchus mykiss
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Euteleostei;
               Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE      1 (bases 1 to 1084)
AUTHORS        Wang, T. and Secombes, C.J.
TITLE          Cloning, expression and phylogenetic analysis of rainbow trout beta
               thymosin
JOURNAL        Unpublished
                2 (bases 1 to 1084)
AUTHORS        Wang, T.
TITLE          Direct Submission
JOURNAL        Submitted (06-OCT-1999) Wang T., Department of Zoology, University
                of Aberdeen, Tillydrone Avenue, Aberdeen, AB24 2TZ, UNITED KINGDOM
FEATURES       Location/Qualifiers
               1..1084
               /organism="Oncorhynchus mykiss"
               /db_xref="taxon:8022"
               /clone="af11"
               /cell_type="leukocyte"
               /tissue_type="head kidney"
               /clone_lib="lambda ZAPK"
               /country="United Kingdom"
               1..56
               57..188
               /codon_start=1
               /product="beta thymosin"
               /protein_id="CAC12827.1"
               /db_xref="GI:10798668"
               /translation="MSDKPNLEEVASFDTKLKKTETCEKNPLPTKETIEQKQASS"
               189..1026
               polyA_signal 1027..1032
BASE COUNT     331 a 241 c 220 g 292 t
ORIGIN

Query Match      4.3%; Score 35; DB 5; Length 1084;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;

misc_feature     /note="assembly_fragment"
5548..6776
misc_feature     /note="assembly_fragment"
6877..8215
misc_feature     /note="assembly_fragment"
8316..9961
misc_feature     /note="assembly_fragment"
10062..12224
misc_feature     /note="assembly_fragment"
12325..13794
misc_feature     /note="assembly_fragment"
13895..16857
misc_feature     /note="assembly_fragment"
16958..18266
misc_feature     /note="assembly_fragment"
18367..20190
misc_feature     /note="assembly_fragment"
20291..21460

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 775 TTTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
|||||...|||||
Db 1040 TTTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1074

RESULT 7
AC073453       61373 bp      DNA      linear      HTG 18-JUN-2000
LOCUS          Homo sapiens clone RP11-772N1, LOW-PASS SEQUENCE SAMPLING.
DEFINITION     AC073453
ACCESSION      AC073453.1 GI:8571751
VERSION        HTG; HTGS_PHASE0.
KEYWORDS       Homo sapiens.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 61373)
AUTHORS        Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
               Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
               Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
               Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
               Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
               Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
               Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
               Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
               Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
               Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehotzky, J.,
               Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
               McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
               Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
               Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
               O'Neill, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N.,
               Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
               Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
               Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
               Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
               Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
               Young, G., Zainoun, J., Zimmer, A. and Zody, M.
               Direct Submission
               Submitted (18-JUN-2000) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
               All repeats were identified using RepeatMasker:
               Smit, A.F.A. & Green, P. (1996-1997)
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ----- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
               ----- Project Information
               Center project name: L8647
               Center clone name: 772_N_1
               -----
               * NOTE: This record contains 79 individual
               * sequencing reads that have not been assembled into
               * contigs. Runs of N are used to separate the reads
               * and the order in which they appear is completely
               * arbitrary. Low-pass sequence sampling is useful for
               * identifying clones that may be gene-rich and allows
               * overlap relationships among clones to be deduced.
               * However, it should not be assumed that this clone
               * will be sequenced to completion. In the event that
               * the record is updated, the accession number will
               * be preserved.
               * 1 683: contig of 683 bp in length
               * 684 783: gap of 100 bp
               * 784 1456: contig of 673 bp in length

```





## RESULT 8

AC021344/c  
 LOCUS AC021344  
 DEFINITION Homo sapiens clone RP11-24J4, LOW-PASS SEQUENCE SAMPLING.  
 AC021344  
 VERSION AC021344.2 GI:9130794  
 KEYWORDS HTG: HTGS\_PHASE0.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

## REFERENCE

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
 TITLE 1 (bases 1 to 72632)  
 JOURNAL Mammalia: Eutheria: Primates: Catarrhini; Homnidae; Homo.  
 REFERENCE Unpublished  
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

## TITLE

JOURNAL Direct Submission  
 COMMENT Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jul 13, 2000 this sequence version replaced gi:6705762.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L4482

Center clone name: 24\_J\_4

\* NOTE: This record contains 74 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1  
 884 983: gap of 100 bp  
 984 1863: contig of 880 bp in length.  
 1864 1963: gap of 100 bp  
 1964 2847: contig of 884 bp in length.  
 2848 2947: gap of 100 bp  
 2948 3826: contig of 879 bp in length.  
 3827 3926: gap of 100 bp  
 3927 4800: contig of 874 bp in length.  
 4801 4900: gap of 100 bp  
 4901 5756: contig of 856 bp in length.  
 5757 5856: gap of 100 bp  
 5857 6700: contig of 844 bp in length

\* 6701 6800: gap of 100 bp  
 \* 6801 7650: contig of 850 bp in length  
 \* 7651 7750: gap of 100 bp  
 \* 7751 8647: contig of 897 bp in length  
 \* 8648 8747: gap of 100 bp  
 \* 8748 9632: contig of 885 bp in length  
 \* 9633 9732: gap of 100 bp  
 \* 9733 10615: contig of 883 bp in length  
 \* 10616 10715: gap of 100 bp  
 \* 10716 11611: contig of 896 bp in length  
 \* 11612 11711: gap of 100 bp  
 \* 11712 12588: contig of 877 bp in length  
 \* 12589 12688: gap of 100 bp  
 \* 12689 13584: contig of 896 bp in length  
 \* 13585 13684: gap of 100 bp  
 \* 13685 14556: contig of 872 bp in length  
 \* 14557 14656: gap of 100 bp  
 \* 14657 15550: contig of 894 bp in length  
 \* 15551 15650: gap of 100 bp  
 \* 15651 16529: contig of 879 bp in length  
 \* 16530 16629: gap of 100 bp  
 \* 16630 17499: contig of 870 bp in length  
 \* 17500 17599: gap of 100 bp  
 \* 17600 18491: contig of 892 bp in length  
 \* 18492 18591: gap of 100 bp  
 \* 18592 19476: contig of 885 bp in length  
 \* 19477 19576: gap of 100 bp  
 \* 19577 20469: contig of 893 bp in length  
 \* 20470 20569: gap of 100 bp  
 \* 20570 21451: contig of 882 bp in length  
 \* 21452 21551: gap of 100 bp  
 \* 21552 22424: contig of 873 bp in length  
 \* 22425 22524: gap of 100 bp  
 \* 22525 23393: contig of 869 bp in length  
 \* 23394 23493: gap of 100 bp  
 \* 23494 24369: contig of 876 bp in length  
 \* 24370 24469: gap of 100 bp  
 \* 24470 25347: contig of 878 bp in length  
 \* 25348 25447: gap of 100 bp  
 \* 25448 26321: contig of 874 bp in length  
 \* 26322 26421: gap of 100 bp  
 \* 26422 27276: contig of 855 bp in length  
 \* 27277 27376: gap of 100 bp  
 \* 27377 28233: contig of 857 bp in length  
 \* 28234 28333: gap of 100 bp  
 \* 28334 29226: contig of 893 bp in length  
 \* 29227 29326: gap of 100 bp  
 \* 29327 30241: contig of 915 bp in length  
 \* 30242 30341: gap of 100 bp  
 \* 30342 31214: contig of 873 bp in length  
 \* 31215 31314: gap of 100 bp  
 \* 31315 32209: contig of 895 bp in length  
 \* 32210 32309: gap of 100 bp  
 \* 32310 33194: contig of 885 bp in length  
 \* 33195 33294: gap of 100 bp  
 \* 33295 34187: contig of 893 bp in length  
 \* 34188 34287: gap of 100 bp  
 \* 34288 35133: contig of 846 bp in length  
 \* 35134 35233: gap of 100 bp  
 \* 35234 36127: contig of 894 bp in length  
 \* 36128 36227: gap of 100 bp  
 \* 36228 37130: contig of 903 bp in length  
 \* 37131 37230: gap of 100 bp  
 \* 37231 38121: contig of 891 bp in length  
 \* 38122 38221: gap of 100 bp  
 \* 38222 39116: contig of 895 bp in length  
 \* 39117 39216: gap of 100 bp  
 \* 39217 40083: contig of 867 bp in length  
 \* 40084 40183: gap of 100 bp  
 \* 40184 41097: contig of 914 bp in length  
 \* 41098 41197: gap of 100 bp  
 \* 41198 42118: contig of 921 bp in length  
 \* 42119 42218: gap of 100 bp

```
* 42219 43105: contig of 887 bp in length
* 43106 43205: gap of 100 bp
* 43206 44101: contig of 896 bp in length
* 44102 44201: gap of 100 bp
* 44202 45069: contig of 888 bp in length
* 45070 45169: gap of 100 bp
* 45170 46041: contig of 872 bp in length
* 46042 46141: gap of 100 bp
* 46142 47006: contig of 865 bp in length
* 47007 47106: gap of 100 bp
* 47107 47962: contig of 856 bp in length
* 47963 48062: gap of 100 bp
* 48063 48954: contig of 892 bp in length
* 48955 49054: gap of 100 bp
* 49055 49931: contig of 877 bp in length
* 49932 50031: gap of 100 bp
* 50032 50912: contig of 881 bp in length
* 50913 51012: gap of 100 bp
* 51013 51907: contig of 895 bp in length
* 51908 52007: gap of 100 bp
* 52008 52882: contig of 875 bp in length
* 52883 52982: gap of 100 bp
* 52983 53879: contig of 897 bp in length
* 53880 53979: gap of 100 bp
* 53980 54874: contig of 895 bp in length
* 54875 54974: gap of 100 bp
* 54975 55837: contig of 863 bp in length
* 55838 55937: gap of 100 bp
* 55938 56860: contig of 923 bp in length
* 56861 56960: gap of 100 bp
* 56961 57801: contig of 841 bp in length
* 57802 57901: gap of 100 bp
* 57902 58759: contig of 858 bp in length
* 58760 58859: gap of 100 bp
* 58860 59749: contig of 890 bp in length
* 59750 59849: gap of 100 bp
* 59850 60742: contig of 893 bp in length
* 60743 60842: gap of 100 bp
* 60843 61738: contig of 896 bp in length
* 61739 61838: gap of 100 bp
* 61839 62728: contig of 890 bp in length
* 62729 62828: gap of 100 bp
* 62829 63704: contig of 876 bp in length
* 63705 63804: gap of 100 bp
* 63805 64694: contig of 890 bp in length
* 64695 64794: gap of 100 bp
* 64795 65702: contig of 908 bp in length
* 65703 65802: gap of 100 bp
* 65803 66688: contig of 886 bp in length
* 66689 66788: gap of 100 bp
* 66789 67681: contig of 893 bp in length
* 67682 67781: gap of 100 bp
* 67782 68673: contig of 892 bp in length
* 68674 68773: gap of 100 bp
* 68774 69666: contig of 893 bp in length
* 69667 69766: gap of 100 bp
```

```
Query Match 4.3%; Score 35; DB 2: Length 72632;
Best Local Similarity 100.0%; Pred. No. 2.9e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 774 CTTTGTCTTAAAAA.....AAAAAAAAAAAAA 808
|||||.....
Db 56858 CTTTGTCTTAAAAA.....AAAAAAAAAAAAA 56824
```

```
RESULT 9
AC009814/c AC009814 linear HTG 26-MAY-2000
LOCUS Homo sapiens clone RP11-115G17, WORKING DRAFT SEQUENCE, 30
DEFINITION unordered pieces.
ACCESSION AC009814
VERSION AC009814.3 GI:8072479
```

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 160197)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-115G17  
Unpublished  
2 (bases 1 to 160197)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferrel, P., FitzHugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,  
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,  
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mchaleckyj, J.,  
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (02-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:6088005.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RN/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1866  
Center clone name: L15-G.17  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 103678 bases at least Q40  
Consensus quality: 127132 bases at least Q30  
Consensus quality: 142688 bases at least Q20  
Insert size: 161000; agarose-fp  
Quality coverage: 3.3 in Q20 bases; agarose-fp  
Quality coverage: 3.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 30 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1046: contig of 1046 bp in length  
\* 1047 1146: gap of 100 bp  
\* 1147 2209: contig of 1063 bp in length  
\* 2210 2309: gap of 100 bp  
\* 2310 3410: contig of 1101 bp in length  
\* 3411 3510: gap of 100 bp  
\* 3511 4951: contig of 1441 bp in length  
\* 4952 5051: gap of 100 bp  
\* 5052 6769: contig of 1718 bp in length  
\* 6770 6869: gap of 100 bp  
\* 6870 8564: contig of 1695 bp in length  
\* 8565 8664: gap of 100 bp  
\* 8665 10597: contig of 1933 bp in length





```

BASE COUNT      141 a      196 c      210 g      174 t      2 others
ORIGIN

Query Match      4.2%; Score 34; DB 9; Length 723;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
      |||...|||
Db 561 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 528

RESULT 12
LOCUS AF230276
DEFINITION Triphysaria versicolor clone TvEXP3 alpha-expansin 3 mRNA, complete cds.
ACCESSION AF230276
VERSION AF230276.2 GI:7580482
KEYWORDS
SOURCE Triphysaria versicolor.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Lamiales; Orobanchaceae; Triphysaria.
REFERENCE
AUTHORS Wrobel,R.L. and Voder,J.I.
TITLE Differential RNA expression of alpha-expansin gene family members
in the parasitic angiosperm Triphysaria versicolor
JOURNAL Gene 266 (1-2), 85-93 (2001)
MEDLINE 21186069
PUBMED 11290422
AUTHORS Wrobel,R.L. and Voder,J.I.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2000) Vegetable Crops, University of California, Davis, Davis, CA 95616, USA
REFERENCE
AUTHORS Wrobel,R.L. and Voder,J.I.
TITLE Direct Submission
JOURNAL Submitted (17-APR-2000) Vegetable Crops, University of California, Davis, Davis, CA 95616, USA
REMARK Sequence update by submitter
COMMENT On Apr 17, 2000 this sequence version replaced gi:6942321.
FEATURES
source
Location/Qualifiers
1..1263
/db_xref="taxon:64093"
/clone="TvEXP3"
/tissue_type="roots treated with DMBQ"
/codon_start=1
/evidence=not experimental
/product="alpha-expansin 3"
/protein_id="AAF32409.1"
/db_xref="GI:6942322"
/translacion="MAPLGLFLGLFTLLSOVHVGWGINAHATFYGSDASGTMGGA
CGYNLYSOQYNTALSTAMNTGLSCGYEIRCVNDGKWLPGSLIVATNFCP
PNSALPNAGGWCNPLPHFDLAQPVFOHTAOKAGIVPVAYRRVACRRGGIRFTIN
GHSVENLYVTNVGAGDVHVASIKGSRVWQGNWQSNLNGQSLSFKVT
TSDGRVLSYNVAPARWSFGQTFAQGFR"
BASE COUNT      346 a      227 c      314 g      376 t
ORIGIN

Query Match      4.2%; Score 34; DB 8; Length 1263;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
      |||...|||
Db 1230 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1263

RESULT 13
LOCUS HSM803416
DEFINITION Homo sapiens mRNA; cDNA DKFp313H118 (from clone DKFp313H118).
ACCESSION AL832109
VERSION AL832109.1 GI:21732652
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amlid,C. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFp313H118) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
source
Location/Qualifiers
1..4609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFp313H118"
/tissue_type="cDNA-collection"
/clone_lib="313 (synonym: hlcc2). Vector pTriplex2; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
polyA_signal
polyA_site 4581
BASE COUNT      1206 a      966 c      803 g      1634 t
ORIGIN

Query Match      4.2%; Score 34; DB 9; Length 4609;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
      |||...|||
Db 4574 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 4607

RESULT 14
LOCUS AC004927/c
DEFINITION Homo sapiens PAC clone RP5-910117 from TqII.21-q11.23, complete sequence.
ACCESSION AC004927
VERSION AC004927.2 GI:4309815
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99053792
PUBMED 9847074
REFERENCE
AUTHORS Edwards,J., Lacy,M. and Willson,C.
TITLE The sequence of Homo sapiens PAC clone RP5-910117
JOURNAL Unpublished
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission

```

## JOURNAL

Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 80105)

REFERENCE  
AUTHORS

Waterston,R.

JOURNAL  
TITLE

Submitted (25-FEB-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
5 (bases 1 to 80105)

REFERENCE  
AUTHORS

Waterston,R.

JOURNAL  
TITLE

Submitted (21-DEC-1999) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Mar 1, 1999 this sequence version replaced gi:3213070.

## COMMENT

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: H\_DJ0910117

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send  
<mailto:leegreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by  
Pietler de Jong and coworkers at the Roswell Park Cancer Institute  
(<http://bacpac.med.buffalo.edu>) using the method described by  
Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from  
one male donor.  
The clone may be obtained either from Genome Systems, Inc.  
(<http://www.genomesystems.com>) or Research Genetics, Inc.  
(<http://www.resgen.com>); or from Pietler de Jong.  
VECTOR: pCYPAC2

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-765G7, 200 bp overlap.  
Actual start of this clone is at base position 1 of RP5-910117;  
actual end is at 31497 of RP4-765G7.

## FEATURES

## source

1..80105  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q11.21-q11.23"  
/clone="RP5-910117"  
/clone\_lib="RPCI-5"

## misc\_feature

5..1065  
/note="CpG\_island (%GC=79.0, o/e=0.90, #CpGs=169)"

## repeat\_region

61..102  
/rpt\_family="GC\_rich"

## repeat\_region

146..179

repeat\_region  
309..346  
/rpt\_family="polypurine"  
repeat\_region  
522..546  
/rpt\_family="GC\_rich"  
repeat\_region  
836..887  
/rpt\_family="(CCCCG)n"  
repeat\_region  
1150..1231  
/rpt\_family="MIR"  
misc\_feature  
1711..2059  
/note="match to EST A1193767 (NID:g3744976) qe79g01.x1"  
repeat\_region  
2675..2965  
/rpt\_family="Alu"  
repeat\_region  
3880..4159  
/rpt\_family="Alu"  
repeat\_region  
4616..4827  
/rpt\_family="MIR"  
repeat\_region  
5042..5345  
/rpt\_family="Alu"  
repeat\_region  
5442..5765  
/rpt\_family="Alu"  
repeat\_region  
6005..6294  
/rpt\_family="Alu"  
repeat\_region  
8006..8308  
/rpt\_family="Alu"  
repeat\_region  
8369..8581  
/rpt\_family="MIR"  
repeat\_region  
9338..9649  
/rpt\_family="Alu"  
repeat\_region  
10179..10332  
/rpt\_family="(CACAC)n"  
repeat\_region  
11898..12072  
/rpt\_family="MaLR"  
repeat\_region  
12043..12185  
/rpt\_family="MaLR"  
repeat\_region  
13305..13329  
/rpt\_family="AT\_rich"  
repeat\_region  
15481..15589  
/rpt\_family="L2"  
repeat\_region  
15842..15757  
/rpt\_family="L1"  
repeat\_region  
16206..16366  
/rpt\_family="MIR"  
repeat\_region  
16723..16750  
/rpt\_family="AT\_rich"  
repeat\_region  
16751..17038  
/rpt\_family="Alu"  
repeat\_region  
17039..17069  
/rpt\_family="AT\_rich"  
repeat\_region  
17121..17248  
/rpt\_family="Alu"  
repeat\_region  
17674..17952  
/rpt\_family="L2"  
repeat\_region  
17953..18245  
/rpt\_family="Alu"  
repeat\_region  
18379..18681  
/rpt\_family="Alu"  
repeat\_region  
18842..19311  
/rpt\_family="MaLR"  
repeat\_region  
19328..19629  
/rpt\_family="Alu"  
repeat\_region  
20768..20889  
/rpt\_family="MIR"  
repeat\_region  
21047..21141  
/rpt\_family="L2"  
repeat\_region  
21165..21481  
/rpt\_family="Alu"  
repeat\_region  
21940..22245  
/rpt\_family="Alu"  
repeat\_region  
23062..23174  
/rpt\_family="MIR"  
repeat\_region  
23257..23369  
/rpt\_family="L2"

```

repeat_region 23627..23950
/rpt_family="Alu"
repeat_region 25242..25318
/rpt_family="MIR"
repeat_region 26494..26561
/rpt_family="MER1_type"
repeat_region 26562..26827
/rpt_family="Alu"
repeat_region 26828..26935
/rpt_family="MER1_type"
repeat_region 27849..27869
/rpt_family="AT-rich"
repeat_region 28511..28768
/rpt_family="MIR"
repeat_region 28947..29181
/rpt_family="MER2_type"
repeat_region 29242..29359
/rpt_family="Alu"
repeat_region 29360..29410
/rpt_family="(TAA)n"
repeat_region 29411..29577
/rpt_family="Alu"
repeat_region 29748..30051
/rpt_family="Alu"
repeat_region 30306..30341
/rpt_family="AT-rich"
repeat_region 31419..31727
/rpt_family="Alu"
repeat_region 33222..33370
/rpt_family="MIR"
repeat_region 33748..33895

```

Query Match 4.2%; Score 34; DB 9; Length 80105;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
|||||
Db 68489 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 68456

```

## RESULT 15

```

AC011461
LOCUS AC011461 100680 bp DNA linear PRI 14-JUL-2002
DEFINITION Homo sapiens chromosome 19 clone CTC-429L19, complete sequence.
ACCESSION AC011461
VERSION AC011461.5 GI:21747438
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100680)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 100680)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 100680)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 14, 2002 this sequence version replaced gi:14269680.
Draft Sequence Produced by DOE Joint Genome Institute

```

www.tgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 100% of Sequence;  
 Estimated Total Number of Errors is 0.

## FEATURES

```

source
1..100680
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTC-429L19"

```

BASE COUNT 23919 a 26581 c 25942 g 24238 t  
 ORIGIN

Query Match 4.2%; Score 34; DB 9; Length 100680;  
 Best Local Similarity 100.0%; Pred. No. 1e-07;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 776 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
|||||
Db 32997 TTGTTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 33030

```

Search completed: January 25, 2003, 17:53:35  
 Job time : 3621 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 13:07:02 ; Search time 261 Seconds  
(without alignments)  
6980.332 Millicn cell updates/sec

Title: US-09-822-080b-1  
Perfect score: 809  
Sequence: 1 gatgaatctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	4.2	17245	22 AAK83897	Human immune/haema
2	33	4.1	2736	24 AAL46316	Rat M30 coding seq
3	33	4.1	8855	22 ABA19146	Human nervous syst
4	33	4.1	32188	22 AAS28365	Genomic sequence #
5	32	4.0	347	22 AAI83215	Human polynucleoti
6	32	4.0	354	23 ABV49061	Human prostate exp
7	32	4.0	400	22 AAI92151	Human polynucleoti
8	32	4.0	623	21 AAI16202	Human colon cancer
9	31	3.8	114	23 AAS57516	CDNA #192 encoding

10	31	3.8	176	22 AAH25484	Nucleotide fragmen
11	31	3.8	202	22 AAL18363	Human breast cance
12	31	3.8	205	24 ABL77170	Human ovarian can
13	31	3.8	211	23 ABV18862	Human prostate exp
14	31	3.8	228	23 ABV49220	Human prostate exp
15	31	3.8	228	24 ABN96972	Gene #3470 used to
16	31	3.8	272	23 ABV17661	Human prostate exp
17	31	3.8	316	23 ABV47962	Human prostate exp
18	31	3.8	336	23 ABV48527	Human prostate exp
19	31	3.8	359	23 ABV56960	Human prostate exp
20	31	3.8	373	22 AAL19814	Human prostate exp
21	31	3.8	407	22 AAI89211	Human breast cance
22	31	3.8	418	22 AAI89274	Human polynucleoti
23	31	3.8	437	20 AAX61357	Human polynucleoti
24	31	3.8	457	22 AAI85461	DNA encoding a hum
25	31	3.8	457	23 AAI85463	Human polynucleoti
26	31	3.8	465	23 ABV18744	Human prostate exp
27	31	3.8	473	24 ABQ58556	Human prostate exp
28	31	3.8	510	23 ABV47455	Human colon cancer
29	31	3.8	641	21 AAZ33356	Human prostate exp
30	31	3.8	650	22 AAH34672	Human prostate exp
31	31	3.8	687	24 ABQ65739	Human secreted pro
32	31	3.8	786	20 AAZ00441	Human colon cancer
33	31	3.8	956	21 AAC77888	Arabidopsis thalia
34	31	3.8	1007	21 AAC56066	Human secreted pro
35	31	3.8	1093	21 AAC74417	Human cancer assoc
36	31	3.8	1093	21 AAC69118	Eucalyptus grandis
37	31	3.8	1107	22 AAH34072	Human secreted pro
38	31	3.8	1172	21 AAC98117	Human secreted pro
39	31	3.8	1172	22 AAH33220	Human colon cancer
40	31	3.8	1315	22 AAD05464	Human colon cancer
41	31	3.8	1392	22 AAF32773	Human secreted pro
42	31	3.8	1435	24 ABA98819	Rice ALS small sub
43	31	3.8	1477	22 AAS02065	Human MANGO 511 cd
44	31	3.8	1477	22 AAS02102	Human MANGO 511, v
45	31	3.8	1477	22 AAS02103	Human MANGO 511, v

## ALIGNMENTS

RESULT 1  
AAK83897  
ID AAK83897 standard; DNA; 17245 BP.  
XX  
AC AAK83897;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38709.  
XX  
DE Human; immune; haematopoietic; Immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.

us-09-822-080b-1.oli.rng

Mon Jan 27 12:25:12 2003

PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000;	2000US-0216880.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	14-JUL-2000;	2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224518.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246611.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226681.	PR	08-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0226686.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	01-DEC-2000;	2000US-0250160.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250391.
PR	12-SEP-2000;	2000US-02321968.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232399.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232400.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251989.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234223.	PR	11-DEC-2000;	2000US-0254097.
PR	21-SEP-2000;	2000US-0234274.	PR	05-JAN-2001;	2001US-0259678.
PR	23-SEP-2000;	2000US-0234597.	XX		
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	PI	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	PI		
PR	29-SEP-2000;	2000US-0236327.	XX	WPI; 2001-483426/52.	
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236368.	XX		
PR	29-SEP-2000;	2000US-0236369.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PR	29-SEP-2000;	2000US-0236370.	PT	useful for preventing, diagnosing and/or treating cancers and	
PR	02-OCT-2000;	2000US-0236802.	PT	metastasis -	
PR	02-OCT-2000;	2000US-0237037.	XX		
PR	02-OCT-2000;	2000US-0237037.	XX	Disclosure: SEQ ID NO 38709; 3071pp + Sequence Listing; English.	
PR	02-OCT-2000;	2000US-0237038.	XX		
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	02-OCT-2000;	2000US-0237040.	CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)	
PR	13-OCT-2000;	2000US-0239935.	CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic	
PR	13-OCT-2000;	2000US-0239937.	CC	activity, and can be used in gene therapy and vaccine production. (I)	
PR	20-OCT-2000;	2000US-0240960.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
PR	20-OCT-2000;	2000US-0241121.	CC	treatment of diseases associated with inappropriate (I) expression. For	
PR	20-OCT-2000;	2000US-0241185.	CC	example, they may be used to treat disorders associated with decreased	
PR	20-OCT-2000;	2000US-0241186.	CC	expression by rectifying mutations or deletions in a patient's genome	
PR	20-OCT-2000;	2000US-0241187.	CC	that affect the activity of (I) by expressing inactive proteins or to	

CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX

SQ Sequence 17245 BP; 4069 A; 4768 C; 4523 G; 3885 T; 0 other;

Query Match 4.2%; Score 34; DB 22; Length 17245;  
 Best Local Similarity 100.0%; Pred. No. 0.00019;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 TTGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
 Db 2964 TTGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2997

## RESULT 2

AA146316  
 ID AAL46316 standard; cDNA; 2736 BP.

XX  
 AC AAL46316;

XX  
 DT 19-JUL-2002 (first entry)

XX  
 DE Rat M30 coding sequence SEQ ID NO: 1.

XX  
 KW Neurodegenerative disease; M30; M31; M32; M33; stroke;  
 KW fragile X syndrome; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; multiple sclerosis; ovarian cancer;  
 KW neurodegeneration; immune disorder; autoimmune disease; allergy;  
 KW infection; leukaemia; inflammation; neuroprotective; cerebroprotective;  
 KW immunosuppressive; cytostatic; nontropic; antiparkinsonian; anti-allergic;  
 KW virucide; anti-inflammatory; gene; ss.

XX  
 OS Rattus sp.

XX  
 PN WO200221138-A2.

XX  
 PD 14-MAR-2002.

XX  
 PF 07-SEP-2001; 2001WO-EP10366.

XX  
 PR 07-SEP-2000; 2000US-0657479.

XX  
 PA (AXAR-) AXARON BIOSCIENCE AG.

XX  
 PI Schneider A, Hiemisch H, Rossner M, Klugmann M, Naim J;  
 PI Eisenhardt G, Kuner R, Lanahan A, Worley P, Spielvogel D;  
 PI Scheek S;

DR  
 DR WPI; 2002-292287/33.

DR  
 DR P-PSDB; AAO17515.

XX  
 PT Diagnosis of neurodegenerative disease comprises detecting level of  
 PT M30-family proteins.

XX  
 PS Claim 3; Page 75-78; 130pp; German.

XX  
 CC The present invention relates to a method of diagnosing neurodegenerative  
 CC diseases, comprising determining the concentration of a protein in a body  
 CC sample, where the protein may be M30 or a variant thereof, M31, M32 or  
 CC M33. The method is used to diagnose neurodegenerative diseases,  
 CC particularly stroke but also e.g. fragile X syndrome, Huntington's,  
 CC Parkinson's and Alzheimer's diseases, multiple sclerosis etc. Also  
 CC overexpression of M31 can be used for diagnosis of carcinoma and sarcoma,  
 CC especially ovarian cancer. The proteins can be used to identify specific  
 CC ligands, potentially useful for treating neurodegeneration, immune-system  
 CC disorders (e.g. autoimmune diseases, allergy, viral infection,

CC leukaemia, inflammation etc.), carcinoma and sarcoma. Inhibitors of the  
 CC interaction between the proteins and the protein kinase IRAK-1 can be  
 CC used to treat neurodegeneration. The present sequence is a coding  
 CC sequence of a protein used in the method of the invention.

XX  
 SQ Sequence 2736 BP; 757 A; 596 C; 596 G; 787 T; 0 other;

Query Match 4.1%; Score 33; DB 24; Length 2736;  
 Best Local Similarity 100.0%; Pred. No. 0.00067;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809

Db 2697 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 2729

## RESULT 3

ABA19146

ID ABA19146 standard; DNA; 8855 BP.

XX  
 AC ABA19146;

XX  
 DT 23-JAN-2002 (first entry)

XX  
 DE Human nervous system related polynucleotide SEQ ID NO 11477.

XX  
 KW Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW anti-allergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX  
 OS Homo sapiens.

XX  
 PN WO200159063-A2.

XX  
 PD 16-AUG-2001.

XX  
 PF 17-JAN-2001; 2001WO-US01334.

XX  
 PR 31-JAN-2000; 2000US-0179065.

XX  
 PR 04-FEB-2000; 2000US-0180628.

XX  
 PR 24-FEB-2000; 2000US-0184664.

XX  
 PR 02-MAR-2000; 2000US-0186350.

XX  
 PR 16-MAR-2000; 2000US-0189874.

XX  
 PR 17-MAR-2000; 2000US-0190076.

XX  
 PR 18-APR-2000; 2000US-0198123.

XX  
 PR 19-MAY-2000; 2000US-0205515.

XX  
 PR 07-JUN-2000; 2000US-0209467.

XX  
 PR 28-JUN-2000; 2000US-0214886.

XX  
 PR 30-JUN-2000; 2000US-0215135.

XX  
 PR 07-JUL-2000; 2000US-0216647.

XX  
 PR 07-JUL-2000; 2000US-0216880.

XX  
 PR 11-JUL-2000; 2000US-0217487.

XX  
 PR 11-JUL-2000; 2000US-0217496.

XX  
 PR 14-JUL-2000; 2000US-0218290.

XX  
 PR 26-JUL-2000; 2000US-0220963.

XX  
 PR 26-JUL-2000; 2000US-0220964.

XX  
 PR 14-AUG-2000; 2000US-0224518.

XX  
 PR 14-AUG-2000; 2000US-0224519.

XX  
 PR 14-AUG-2000; 2000US-0225213.

XX  
 PR 14-AUG-2000; 2000US-0225214.

XX  
 PR 14-AUG-2000; 2000US-0225266.

XX  
 PR 14-AUG-2000; 2000US-0225267.

XX  
 PR 14-AUG-2000; 2000US-0225268.

XX  
 PR 14-AUG-2000; 2000US-0225270.

XX  
 PR 14-AUG-2000; 2000US-0225447.

XX  
 PR 14-AUG-2000; 2000US-0225757.

XX  
 PR 14-AUG-2000; 2000US-0225758.

XX  
 PR 14-AUG-2000; 2000US-0225759.

XX  
 PR 18-AUG-2000; 2000US-0226279.

PR	17-NOV-2000;	2000US-0249207.	
PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249244.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249245.	
PR	17-NOV-2000;	2000US-0249264.	
PR	17-NOV-2000;	2000US-0249265.	
PR	17-NOV-2000;	2000US-0249297.	
PR	17-NOV-2000;	2000US-0249299.	
PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250391.	
PR	01-DEC-2000;	2000US-0251160.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-02559678.	
XX			
XX	(HUMA- ) HUMAN GENOME SCI INC.		
XX			
XX	Rosen CA, Barash SC, Ruben SM;		
PI	WPI; 2001-541565/60.		
XX			
DR	Nucleic acids encoding 3224 human nervous system antigen polypeptides,		
PT	useful for preventing, diagnosing and/or treating nervous system		
PT	cancers and metastases -		
PT			
XX	Disclosure; SEQ ID NO 11477; 1701pp + Sequence Listing; English.		
XX			
XX	The invention relates to novel genes (ABA11004-ABA21534) and proteins		
CC	(ABB14678-ABB18001) useful for preventing, treating or ameliorating		
CC	medical conditions e.g. by protein or gene therapy. The genes are		
CC	isolated from a range of human tissues disclosed in the specification.		
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful		
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast		
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone		
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;		
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune		
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's		
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative		
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;		
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and		
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal		
CC	and parasitic infections.		
CC	Note: The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence 8855 BP; 2512 A; 1891 C; 2192 G; 2259 T; 1 other;		
	Query Match 4.1%; Score 33; DB 22; Length 8855;		
	Best Local Similarity 100.0%; Pred. No. 0.00054;		
	Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps		
QY	777 TGTTCCTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 809		
Db	2645 TGTTCCTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 2677		

Query Match	4.1%	Score 33;	DB 22;	Length 8855;
Best Local Similarity	100.0%;	Pred. No. 0.00554;		
Matches 33;	Conservative 0;	Mismatches 0;	Indels	
QY 777	TGTTTCTAAAAA	AAAAAAAAAAAAAAAAAAAA	AAAAA	809
Db 2645	TGTTTCTAAAAA	AAAAAAAAAAAAAAAAAAAA	AAAAA	2677

RESULT 4  
AAS28365  
ID AAS28365 standard; DNA; 32188 BP.  
XX AC AAS28365;  
XX DT 07-NOV-2001 (first entry)  
XX DE Genomic sequence #205 encoding for novel human respiratory antigen.  
XX KW Human; respiratory antigen; respiratory disorder; throat disorder;  
XX KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;  
XX KW anti allergic; anti asthmatic; anti inflammatory; olfactory;  
XX KW respiratory active; ds.  
OS Homo sapiens.  
XX WO20015448-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01333.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

```

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-476224/51.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure: SED ID No 799; 546pp; English.
XX
XX The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 32188 BP; 8564 A; 6422 C; 6550 G; 10652 T; 0 other;
SQ

Query Match 4.1%; Score 33; DB 22; Length 32188;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 15956 TGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 15988

RESULT 5
AAI83215
ID AAI83215 standard; cDNA; 347 BP.
XX
XX AAI83215;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3275.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
PN

17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-476224/51.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure: SED ID No 799; 546pp; English.
XX
XX The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 32188 BP; 8564 A; 6422 C; 6550 G; 10652 T; 0 other;
SQ

Query Match 4.1%; Score 33; DB 22; Length 32188;
Best Local Similarity 100.0%; Pred. No. 0.00042;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 15956 TGTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 15988

RESULT 5
AAI83215
ID AAI83215 standard; cDNA; 347 BP.
XX
XX AAI83215;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3275.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
PN

17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-476224/51.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure: SED ID No 799; 546pp; English.
XX
XX The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 32188 BP; 8564 A; 6422 C; 6550 G; 10652 T; 0 other;
SQ

Query Match 4.0%; Score 32; DB 22; Length 347;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 778 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 60 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 91

RESULT 6
ABV49061
ID ABV49061 standard; cDNA; 354 BP.
XX
XX ABV49061;
XX
XX 17-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 49052.
DE
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
XX 16-NAR-2000; 2000US-189862P.
XX 25-MAY-2000; 2000US-207454P.
XX 09-JUN-2000; 2000US-211314P.
XX 18-JUL-2000; 2000US-219007P.
XX

```

```

PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 9597; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for:
XX (a) assessing whether a patient is afflicted with prostate cancer;
XX (b) monitoring the progression of prostate cancer in a patient;
XX (c) assessing the efficacy of a test compound to inhibit prostate
XX cancer in a patient;
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound;
XX (g) determining whether prostate cancer has metastasized in a patient;
XX (h) assessing the aggressiveness or indolence of prostate cancer in a
XX patient;
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 354 BP; 100 A; 108 C; 73 G; 73 T; 0 other;
XX
XX Query Match 4.0%; Score 32; DB 23; Length 354;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 778 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
XX
XX Db 259 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 290
XX
XX RESULT 7
XX AA192151
XX ID AA192151 standard; cDNA; 400 BP.
XX
XX AC AA192151;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 12211.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AA012220.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 12211; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA00010-AA013910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 400 BP; 176 A; 51 C; 63 G; 106 T; 4 other;
XX
XX Query Match 4.0%; Score 32; DB 22; Length 400;
XX Best Local Similarity 100.0%; Pred. No. 0.0024;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 778 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
XX
XX Db 284 GTTCTTAAAAA...AAAAAAAAAAAAAAAAAAAA 315
XX
XX RESULT 8
XX AA16202/c
XX ID AA16202 standard; DNA; 623 BP.
XX
XX AC AA16202;
XX
XX 14-JUN-2000 (first entry)
XX
XX Human colon cancer differentially expressed nucleotide sequence #207.
XX
XX Colon cancer; detect; differential expression; human; treatment;
XX detect mutation; non-invasive diagnostic method; ds.
XX
XX Homo sapiens.
XX
XX WO200012702-A2.
XX
XX 09-MAR-2000.
XX
XX 30-AUG-1999; 99WO-US19424.
XX
XX 31-AUG-1998; 98US-0098639.
XX
XX 27-JAN-1999; 99US-0117393.
XX
XX (FARB ) BAYER CORP.
XX
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;
XX Catino TJ, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;
XX Schlegel R;
XX
XX WPI; 2000-256641/22.
XX
XX Novel nucleic acids and proteins for identifying therapeutic agents
XX useful for treating and diagnosing cancer, especially colon cancer -
XX
XX Claim 16; Page 206-207; 345pp; English.
XX
XX This sequence represents a human nucleotide sequence which is
XX differentially expressed in colon cancer cells compared to the expression
XX levels in normal cells. The nucleotide sequence can be used as a source
XX

```

CC of primers and probes. The nucleotide sequence is useful for determining  
CC the phenotype of a cell by detecting the differential expression of the  
CC sequence relative to a normal cell. The probes derived from the sequence  
CC can also be used to determine the phenotype of cells in a sample. Probes  
CC and antibodies which hybridise to the nucleotide sequence can also be  
CC used to determine the phenotype of a cell. The primers are useful for  
CC detecting a mutation in a test nucleotide sequence and also for detecting  
CC cancer, preferably colon cancer. Antibodies against the protein encoded  
CC by the nucleotide sequence can also be used in a method to detect colon  
CC cancer. The diagnostic method is non-invasive and accurate for diagnosing  
CC colon cancer at an early stage.

XX SQ Sequence 623 BP; 177 A; 106 C; 144 G; 187 T; 9 other;

Query Match 4.0%; Score 32; DB 21; Length 623;  
Best Local Similarity 100.0%; Pred. No. 0.0022;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 778 GTTCTAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 38 GTTCTAATAAAAAAAAAAAAAAAAAAAAAA 7

RESULT 9  
AAS57516  
ID AAS57516 standard; cDNA; 114 BP.

XX AC AAS57516;

XX 13-FEB-2002 (first entry)

XX cDNA #192 encoding portion of a human colon tumour protein.

XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.

XX Homo sapiens.

XX WQ200173027-A2.

XX 04-OCT-2001.

XX 22-MAR-2001; 2001WO-US09246.

XX 24-MAR-2000; 2000US-191597P.

XX 04-MAY-2000; 2000US-202024P.

XX 03-MAY-2000; 2000US-202189P.

XX (CORI-) CORIXA CORP.

XX Meagher MJ, Xu J, King GE;

XX WPI; 2001-611627/70.

XX New colon tumour proteins and related nucleic acid, useful for  
XX treatment, prevention, diagnosis and monitoring of cancer -  
XX Claim 4; Page 90; 299pp; English.

XX Th present invention relates to the isolation of novel cDNA sequences  
XX encoding for at least an immunogenic portion of human colon tumour  
XX proteins. The sequences of the invention are useful in pharmaceutical  
XX compositions and vaccines for the prevention and treatment of cancers  
XX such as colon cancer. They are also useful for the diagnosis and  
XX monitoring of such cancers. Antibodies to the colon tumour proteins  
XX and antigen presenting cells that express polynucleotides encoding  
XX colon tumour proteins can be used to inhibit the development of  
XX cancers. T-cells that react specifically with colon tumour proteins  
XX are useful for removing tumour cells from samples (e.g. blood) and  
XX for cancer treatment. The polynucleotides sequences are also useful in  
XX gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the  
XX invention that encode for portions of human colon tumour proteins.

XX SQ Sequence 114 BP; 64 A; 6 C; 9 G; 35 T; 0 other;

Query Match 3.8%; Score 31; DB 23; Length 114;  
Best Local Similarity 100.0%; Pred. No. 0.0074;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 80 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 110

RESULT 10  
AAH25484  
ID AAH25484 standard; DNA; 176 BP.

XX AC AAH25484;

XX 22-AUG-2001 (first entry)

XX Nucleotide fragment of ribulose-1,5-bisphosphate carboxylase cDNA.

XX Rubisco; small subunit gene promoter; storage reserve; seed;  
XX transgenic plant; ribulose-1,5-bisphosphate carboxylase; ss.

XX Brassica campestris.

XX WO200141559-A1.

XX 14-JUN-2001.

XX 08-DEC-2000; 2000WO-FI01081.

XX 10-DEC-1999; 99FI-0002659.

XX (UNIC-) UNICROP LTD.

XX Kuvshinov V, Kanerva A, Koivu K, Pehu E;

XX WPI; 2001-381420/40.

XX Novel process of converting storage reserves of dicot seeds into  
XX compositions comprising desired gene products, based on source-sink  
XX principle -

XX Example 2; Page 28; 54pp; English.

XX The present sequence represents a fragment of ribulose-1,5-bisphosphate  
XX carboxylase (rubisco) cDNA. The rubisco promoter is used in the process  
XX of the invention. The specification describes a process for converting  
XX storage reserves (such as protein, carbohydrate and lipid reserves) in  
XX dicot plant seeds into compositions of dicot seeds into compositions  
XX comprising one or more desired gene products, based on a source-sink  
XX principle. The process comprises harnessing the regulatory sequences  
XX of transient proteins accumulating during the initiation of  
XX germination for the production of desired gene products. The method  
XX provides a more feasible, cost-effective, environmentally friendly  
XX process and production system for producing gene products, especially  
XX proteinaceous gene products in the cotyledons of transgenic dicot seeds.

XX SQ Sequence 176 BP; 73 A; 21 C; 15 G; 67 T; 0 other;

Query Match 3.8%; Score 31; DB 22; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 132 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 162

RESULT 11  
AAL18363/c  
ID AAL18363 standard; cDNA; 202 BP.

Query Match 3.8%; Score 31; DB 22; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.0068;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 809  
|||||  
Db 132 TTCTTAATAAAAAAAAAAAAAAAAAAAAAA 162

RESULT 11  
AAL18363/c  
ID AAL18363 standard; cDNA; 202 BP.



AC AAL18363;  
 XX  
 DT 07-DEC-2001 (first entry)  
 XX  
 DE Human breast cancer expressed polynucleotide 10820.  
 XX  
 KW Human; breast cancer; cell marker; cytostatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200151628-A2.  
 XX  
 PD 19-JUL-2001.  
 XX  
 PF 10-JAN-2001; 2001WO-US00798.  
 XX  
 PR 14-JAN-2000; 2000US-0176077.  
 PR 14-MAR-2000; 2000US-0189167.  
 PR 24-MAR-2000; 2000US-0192099.  
 PR 29-MAR-2000; 2000US-0193480.  
 PR 15-MAY-2000; 2000US-0205230.  
 PR 09-JUN-2000; 2000US-0211315.  
 PR 25-JUL-2000; 2000US-0220534.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Lillie J, Xu Y, Wang Y, Steinmann K;  
 XX  
 DR WPT; 2001-451856/48.  
 XX  
 PT New peptide useful as a marker for the diagnosis of breast cancer -  
 XX  
 PS Claim 1; Page 1926; 3695pp; English.  
 XX  
 CC The invention relates to human breast cancer expressed polynucleotides  
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.  
 XX  
 SQ Sequence 202 BP; 51 A; 35 C; 49 G; 67 T; 0 other;  
 XX  
 Query Match 3.8%; Score 31; DB 22; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 0.0067;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 779 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
 DB 40 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 10  
 RESULT 12  
 ABL77170/c  
 ID ABL77170 standard; cDNA; 205 BP.  
 XX  
 AC ABL77170;  
 XX  
 DT 17-MAY-2002 (first entry)  
 XX  
 DE Human ovarian cancer related cDNA clone SEQ ID NO:148.  
 XX  
 KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200192581-A2.  
 XX  
 PD 06-DEC-2001.  
 XX

PF 29-MAY-2001; 2001WO-US17756.  
 XX  
 PR 26-MAY-2000; 2000US-207484P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Algate PA, Harlocker SL, Jones R;  
 XX  
 DR WPI; 2002-122075/16.  
 XX  
 CC Composition for therapy and diagnosis of ovarian cancer comprising  
 CC polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 CC polypeptide, antibody specific to polypeptide or T cell expressing  
 CC polypeptide  
 XX  
 PS Claim 1; SEQ ID 148; 489pp; English.  
 XX  
 CC The present invention describes a composition (I) comprising: carriers  
 CC and immunostimulants; and a polypeptide (III) of a ovarian tumour  
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence  
 CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to  
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell  
 CC population of (II), or antigen presenting cells that express (II).  
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to  
 CC (S1) can be used for detecting ovarian cancer in a patient's biological  
 CC sample preferably serum or ovarian tissue. The method comprises  
 CC contacting a biological sample from a patient with (IV), detecting the  
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to  
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the  
 CC patient, where the amount of polynucleotide hybridising to (IV) is  
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising  
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells  
 CC specific for an ovarian tumour protein comprising contacting T cells  
 CC with (III) or (II). (III) is useful in design and preparation of  
 CC ribozyme molecules for inhibiting expression of the tumour polypeptides  
 CC and proteins in tumour cells; and to isolate a full length gene from a  
 CC suitable library e.g., a tumour cDNA library using well known  
 CC techniques.  
 XX  
 SQ Sequence 205 BP; 57 A; 38 C; 30 G; 80 T; 0 other;  
 XX  
 Query Match 3.8%; Score 31; DB 24; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 0.0066;  
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 779 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
 DB 193 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 163  
 RESULT 13  
 ABV18862  
 ID ABV18862 standard; cDNA; 211 BP.  
 XX  
 AC ABV18862;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 18853.  
 XX  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.

PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 3100; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 211 BP; 88 A; 31 C; 29 G; 29 T; 34 other;  
Query Match 3.8%; Score 31; DB 23; Length 211;  
Best Local Similarity 100.0%; Pred. No. 0.0066;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
Db 59 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 89  
RESULT 14  
ABV49220  
ID ABV49220 standard; cDNA; 228 BP.  
XX  
AC ABV49220;  
XX  
XX Human prostate expression marker cDNA 49211.  
DE  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
KW  
XX Homo sapiens.  
OS  
XX WO200160860-A2.  
PN  
XX 23-AUG-2001.  
PD  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 9622; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 228 BP; 78 A; 50 C; 50 G; 50 T; 0 other;  
Query Match 3.8%; Score 31; DB 23; Length 228;  
Best Local Similarity 100.0%; Pred. No. 0.0065;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809  
Db 114 TTTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 144  
RESULT 15  
ABN96972/C  
ID ABN96972 standard; DNA; 228 BP.  
XX  
AC ABN96972;  
XX  
XX 13-AUG-2002 (first entry)  
DT  
DE Gene #3470 used to diagnose liver cancer.  
XX  
XX Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumour; cytostatic; expression profile; disease state;  
KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
OS  
XX Homo sapiens.  
XX  
XX WO200229103-A2.  
PN  
XX 11-APR-2002.  
PD  
XX 02-OCT-2001; 2001WO-US030589.  
PF  
XX 02-OCT-2000; 2000US-237054P.  
PR  
XX (GENE-) GENE LOGIC INC.  
PA  
XX Horne D, Alvares C, Petes-Da-Silva S, Vockley JG;  
PI WPI; 2002-426119/45.  
XX  
XX Diagnosing and detecting the progression of liver cancer,  
PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
PT involves detecting the level of expression of two or more genes in a  
PT liver tissue sample -  
XX

PS Claim 1; SEQ ID NO 3470; 298pp; English.

XX  
CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytostatic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 228 BP; 80 A; 27 C; 32 G; 89 T; 0 other;

Query Match 3.8%; Score 31; DB 24; Length 228;

Best Local Similarity 100.0%; Pred. No. 0.0065;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 779 TTCTTAAAAA 809

Db 38 TTCTTAAAAA 8

Search completed: January 25, 2003, 15:49:17  
Job time : 305 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 13:19:38 : Search time 50 Seconds  
(without alignments)  
7269.105 Million cell updates/sec

Title: US-09-822-080B-1

Perfect score: 809

Sequence: 1 gatgaatcctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 396772 seqs, 224632407 residues

Word size : 0

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	33	4.1	157	10	US-09-924-035A-556
C 2	33	4.1	216	10	US-09-770-696-60
C 3	33	4.1	457	10	US-09-960-352-8568
C 4	33	4.1	32188	10	US-09-764-860-799
C 5	31	3.8	114	10	US-09-815-343-192
C 6	31	3.8	205	10	US-09-867-701-148
C 7	31	3.8	228	10	US-09-880-107-3469
C 8	31	3.8	687	10	US-09-770-149-316
C 9	31	3.8	956	10	US-09-925-301-282
C 10	31	3.8	1172	10	US-09-925-299-127
C 11	31	3.8	1435	10	US-09-732-618-15
C 12	31	3.8	1623	10	US-09-846-573B-4
C 13	31	3.8	1996	10	US-09-925-301-207
C 14	31	3.8	2516	10	US-09-954-456-76
C 15	31	3.8	2516	10	US-09-954-456-1155
C 16	30	3.7	39	9	US-10-208-357-4
C 17	30	3.7	135	9	US-09-954-531-306
C 18	30	3.7	187	10	US-09-867-701-9211
C 19	30	3.7	215	10	US-09-770-696-69

20	30	3.7	261	10	US-09-867-701-9941	Sequence 9941, Ap
21	30	3.7	309	10	US-09-867-701-9236	Sequence 9236, Ap
C 22	30	3.7	365	10	US-09-834-975-497	Sequence 497, Ap
23	30	3.7	404	10	US-09-960-352-4144	Sequence 4144, Ap
24	30	3.7	444	10	US-09-770-444-829	Sequence 829, Ap
C 25	30	3.7	449	9	US-09-933-797-90	Sequence 90, Appl
26	30	3.7	620	10	US-09-925-297-303	Sequence 303, Ap
C 27	30	3.7	726	10	US-09-770-149-127	Sequence 127, Ap
C 28	30	3.7	763	10	US-09-770-445-949	Sequence 949, Ap
C 29	30	3.7	1146	10	US-09-822-830A-487	Sequence 487, Ap
C 30	30	3.7	1177	10	US-09-925-300-514	Sequence 514, Ap
C 31	30	3.7	1236	10	US-09-799-777-119	Sequence 119, Ap
C 32	30	3.7	1518	10	US-09-967-552A-37	Sequence 37, Appl
C 33	30	3.7	1723	10	US-09-925-299-172	Sequence 172, Ap
C 34	30	3.7	1933	9	US-10-067-534-1	Sequence 1, Appl
C 35	30	3.7	1967	10	US-09-925-301-565	Sequence 565, Ap
C 36	30	3.7	2000	9	US-09-938-842A-2828	Sequence 2828, Ap
C 37	30	3.7	2054	10	US-09-729-674-99	Sequence 99, Appl
C 38	30	3.7	2227	10	US-09-917-800A-1646	Sequence 1646, Ap
C 39	30	3.7	3176	10	US-09-925-297-347	Sequence 347, Ap
C 40	30	3.7	3533	9	US-09-984-245-24	Sequence 24, Appl
C 41	30	3.7	4665	10	US-09-759-960-7	Sequence 7, Appl
C 42	30	3.7	5668	10	US-09-925-302-342	Sequence 342, Ap
C 43	30	3.7	17493	10	US-09-804-471A-3	Sequence 3, Appl
C 44	29	3.6	186	10	US-09-924-035A-405	Sequence 405, Ap
C 45	29	3.6	187	10	US-09-770-696-246	Sequence 246, Ap

## ALIGNMENTS

RESULT 1  
US-09-924-035A-556/c  
; Sequence 556, Application US/09924035A  
; Patent No. US20020142319A1  
; GENERAL INFORMATION:  
; APPLICANT: Glach, Jrn  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; FILE REFERENCE: thaliana  
; CURRENT FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/148,784  
; PRIOR FILING DATE: 1999-08-13  
; NUMBER OF SEQ ID NOS: 900  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 556  
; LENGTH: 157  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-924-035A-556

Query Match 4.1%; Score 33; DB 10; Length 157;  
Best Local Similarity 100.0%; Pred. No. 0.00083;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 777 TGTTCCTAAAAA.....AAAAAAAAAAAAAAAAAAAAA 809  
Db 91 TGTTCCTAAAAA.....AAAAAAAAAAAAAAAAAAAAA 59

RESULT 2  
US-09-770-696-60/c  
; Sequence 60, Application US/09770696  
; Patent No. US2001004940A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jörn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.

```

APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kricker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE OF INVENTION: thailandia
FILE REFERENCE: 2031US (PARA-020PRV)
CURRENT APPLICATION NUMBER: US/09/770,696
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,278
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 911
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 60
LENGTH: 216
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(216)
OTHER INFORMATION: n = A,T,C or G
US-09-770-696-60

Query Match          4.1%; Score 33; DB 10; Length 216;
Best Local Similarity 100.0%; Pred. No. 0.00078;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 91 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 59

RESULT 3
US-09-960-352-8568/c
; Sequence 8568, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8568
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 37-LIB2809-011-Q1-EI-B10
US-09-960-352-8568

Query Match          4.1%; Score 33; DB 10; Length 457;
Best Local Similarity 100.0%; Pred. No. 0.00069;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 777 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 112 TGTTCCTAAAAA...AAAAAAAAAAAAAAAAAAAA 80

RESULT 4
US-09-764-860-799

```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-867-701-148

Query Match          3.8%; Score 31; DB 10; Length 205;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
    |||...|||
Db 193 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 163

RESULT 7
US-09-880-107-3469/c
; Sequence 3469, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-W0
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3469
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 W04550
US-09-880-107-3469

Query Match          3.8%; Score 31; DB 10; Length 228;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
    |||...|||
Db 38 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 8

RESULT 8
US-09-770-149-316
; Sequence 316, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorchach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maia
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
```

```
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 316
; LENGTH: 687
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-149-316

Query Match          3.8%; Score 31; DB 10; Length 687;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
    |||...|||
Db 654 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 684

RESULT 9
US-09-925-301-282
; Sequence 282, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 282
; LENGTH: 956
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-282

Query Match          3.8%; Score 31; DB 10; Length 956;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
    |||...|||
Db 913 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 943

RESULT 10
US-09-925-299-127
; Sequence 127, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 127
; LENGTH: 1172
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-127

Query Match
Best Local Similarity 100.0%; Score 31; DB 10; Length 1172;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 1090 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1120

RESULT 11
US-09-732-618-15
; Sequence 15, Application US/09732618
; Patent No. US20010044939A1
; GENERAL INFORMATION:
; APPLICANT: Abell, Lynn
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Omolayo O.
; TITLE OF INVENTION: Small subunit of Plant Acetolactate Synthase
; FILE REFERENCE: BB1435 US NA
; CURRENT APPLICATION NUMBER: US/09/732,618
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/174,437
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Microsoft office 97
; SEQ ID NO 15
; LENGTH: 1435
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-732-618-15

Query Match
Best Local Similarity 100.0%; Score 31; DB 10; Length 1435;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 1371 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1401

RESULT 12
US-09-846-573B-4
; Sequence 4, Application US/09846573B
; Patent No. US20020077467A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Moore, Emma E.
; APPLICANT: Raymond, Fenella
; TITLE OF INVENTION: Mammalian Calcitonin-like Polypeptide-1
; FILE REFERENCE: 97-73C1
; CURRENT APPLICATION NUMBER: US/09/846,573B
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: 09/213,634
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/069,976
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-573B-4

Query Match
Best Local Similarity 100.0%; Score 31; DB 10; Length 1623;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 1578 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1608

RESULT 13
US-09-925-301-207
; Sequence 207, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 207
; LENGTH: 1996
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-207

Query Match
Best Local Similarity 100.0%; Score 31; DB 10; Length 1996;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 809
Db 1932 TTCTCTAAAAA...AAAAAAAAAAAAAAAAAAAA 1962

RESULT 14
US-09-954-456-76
; Sequence 76, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
```



; LENGTH: 2516  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-76

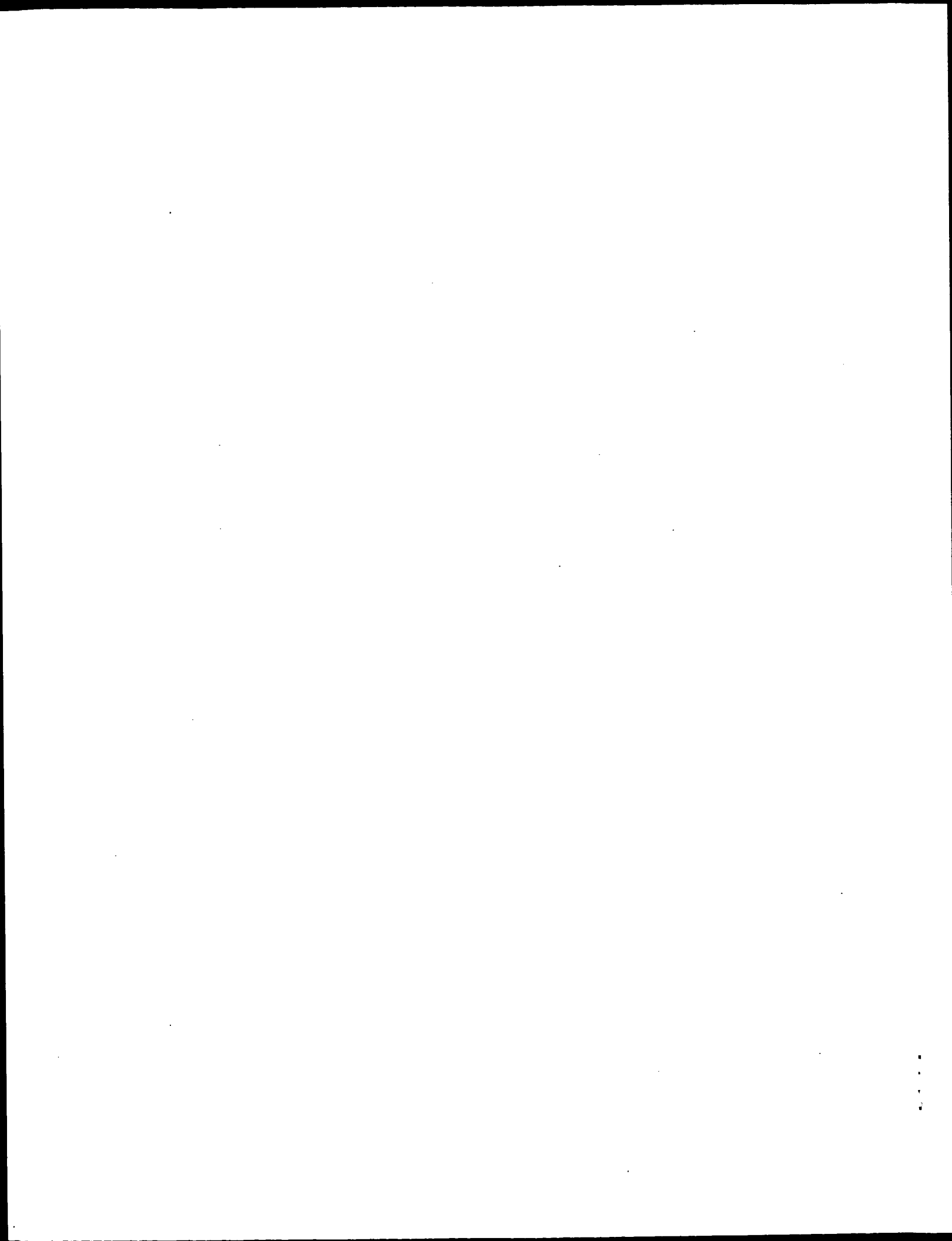
Query Match 3.8%; Score 31; DB 10; Length 2516;  
Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 779 TTCTTAAAAA 809  
Db 2458 TTCTTAAAAA 2488

## RESULT 15

US-09-954-456-1155  
; Sequence 1155, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1155  
; LENGTH: 2516  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-1155

Query Match 3.8%; Score 31; DB 10; Length 2516;  
Best Local Similarity 100.0%; Pred. No. 0.0029;  
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 779 TTCTTAAAAA 809  
Db 2458 TTCTTAAAAA 2488

Search completed: January 25, 2003, 15:51:26  
Job time : 89 secs



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 25, 2003, 10:40:09 ; Search time 2845 seconds  
(without alignments)  
8275.631 Million cell updates/sec

Title: US-09-822-080B-1  
Perfect score: 809  
Sequence: 1 gatgaatctatgttttact.....aaaaaaaaaaaaaaaaaaaaa 809

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.ly.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	783	96.8	783	8	BOU18995	U18995 Brassica ol
2	239.6	29.6	796	8	AY054566	AY054566 Arabidops
3	239	29.5	845	8	AY081323	AY081323 Arabidops
4	239	29.5	70836	8	AC010556	AC010556 Arabidops
5	54.4	6.7	256774	2	AC116964	AC116964 Dictyoste
6	52.6	6.5	133501	2	AC116956	AC116956 Dictyoste
7	52.6	6.5	310779	2	AC005140	AC005140 Plasmodiu
8	52.2	6.5	59004	2	AC115680	AC115680 Dictyoste
9	52	6.4	256172	2	AC005139	AC005139 Plasmodiu
10	51.8	6.4	131665	2	CNS08CB9	AC005139 Plasmodiu
11	51.4	6.4	829	8	AMA012840	AL845347 Oryza sat
12	50.8	6.3	156533	2	AC117070	AJ012840 Avicennia
13	50.4	6.2	1618	3	DDU67089	AJ012840 Avicennia
14	50.4	6.2	163443	2	AC006280	U67089 Dictyoste
15	50.4	6.2	196149	2	AC004709	AC006280 Plasmodiu
16	50.4	6.2	268147	2	AC116966	AC004709 Plasmodiu
17	50.2	6.2	465	8	AF356613	AC116966 Dictyoste
18	50.2	6.2	465	8	AF356614	AF356613 Theobroma
19	50.2	6.2	465	8	AF356618	AF356614 Theobroma
20	50.2	6.2	20043	2	AC116986	AF356618 Theobroma
21	50	6.2	8876	6	AX346979	AC116986 Dictyoste
22	50	6.2	225448	2	AC122296	AX346979 Sequence
23	50	6.2	234488	2	AC109698	AC122296 Mus muscu
24	49.8	6.2	182573	2	AC110879	AC109698 Rattus no
25	49.6	6.1	123280	2	AC117076	AC110879 Oryctolag
26	49.6	6.1	132122	2	AC027327	AC117076 Dictyoste
27	49.6	6.1	184677	2	AC107358	AC027327 Homo sapi
28	49.4	6.1	447	3	AY037520	AC107358 Rattus no
29	49.4	6.1	72072	2	AC111576	AY037520 Tetreure
30	49.2	6.1	845	8	AF128268	AC111576 Rattus no
31	49.2	6.1	8443	2	AC115587	AF128268 Glycine m
32	49.2	6.1	67970	3	PFMAL1P3	AC115587 Dictyoste
33	49.2	6.1	98734	2	PFMAL1P2	AL031746 Plasmodiu
34	49.2	6.1	253756	2	AC093975	AL031746 Plasmodiu
35	49	6.1	83110	2	PFMAL13PB	AC093975 Rattus no
36	49	6.1	157592	2	AC107440	AL109814 Plasmodiu
37	48.8	6.0	8000	3	AF482387	AC107440 Rattus no
38	48.8	6.0	199551	2	AC006281	AF482387 Dictyoste
39	48.6	6.0	465	8	AF356609	AC006281 Plasmodiu
40	48.4	6.0	2979	9	BC031084	AF356609 Theobroma
41	48.4	6.0	12029	3	AE001381	BC031084 Homo sapi
42	48.4	6.0	205429	2	AC005506	AE001381 Plasmodiu
43	48.2	6.0	9958	3	AF246689	AC005506 Plasmodiu
44	48.2	6.0	12500	3	AF482381	AF246689 Dictyoste
45	48.2	6.0	21202	18	PFSC04088	AF482381 Dictyoste
						AL010273 Plasmodiu

# ALIGNMENTS

RESULT 1  
BOU18995  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

BOU18995  
Brassica  
U18995  
U18995.1  
GI:841207  
Brassica oleracea.  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Rosidiales; Magnoliophyta; eudicotyledons; core eudicots;  
1 (bases 1 to 783)  
Williams,D.L., Kain,W.C. and Broadway,R.M.  
Isolation and characterization of a serine proteinase inhibitor

783 bp mRNA linear PLN 04-SEP-1997  
oleracea trypsin inhibitor propeptide mRNA, complete cds.

JOURNAL	CDNA (Accession No. U18995) from cabbage (PCR97-083)
REFERENCE	Plant Physiol. 114, 747 (1997)
AUTHORS	2 (bases 1 to 783)
TITLE	Williams,D.L.
JOURNAL	Direct Submission
FEATURES	Submitted (22-DEC-1994) David L. Williams, NYSAES Cornell University, Geneva, NY 14456, USA Location/Qualifiers 1. .783 /organism="Brassica oleracea" /strain="Superpack" /db_xref="taxon:3712" /clone_lib="lambda zap II" /dev_stage="mature cabbage" 2. .646 /citation=[1] /codon_start=1 /product="trypsin inhibitor propeptide" /protein_id="AAB68964.1" /db_xref="GI:841208" /translation="MNPMPFYLFAETTVLAAATANAGPVLDTDCDIIFDGSYYVLPILW GPVGGGLTVSRGNOCPLFIGOERSEVNRGIPVKFSNRKSRVPEEENINIKMDV EPICAGASAWWVTPAPSRSLFIAGPKPEAGGEDSSRSFQIKTKAKLNAYKFV FCSGNDICIDVGNKEEGVRLVGLSTPFPFAFFVVFVKATGDTSSKTMISII" 2. .64
source	sig_peptide 194 a 186 c 181 g 222 t BASE COUNT ORIGIN
CDS	Query Match 96.8%; Score 783; DB 8; Length 783; Best Local Similarity 100.0%; Pred. No. 3 6e-195; Matches 783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	Qy 1 GATGAATCCTATGTTTACTTCCTTTGCGCTTTACCACTGTTTGGCGCGGACCGCAA 60 Db 1 GATGAATCCTATGTTTACTTCCTTTGCGCTTTACCACTGTTTGGCGCGGACCGCAA 60
	Qy 61 CGCTGCACAGTTCGACACTGATGGTGATATCATATTCGACGGCAGTTACTACGTTCT 120 Db 61 CGCTGGACCAGTTCGACACTGATGGTGATATCATATTCGACGGCAGTTACTACGTTCT 120
	Qy 121 CCCCTCATCTGGGCGCTTACAGTGGCGGCTTAACCTCTCGTCCGTCGTGGCAACCA 180 Db 121 CCCCTCATCTGGGCGCTTACAGTGGCGGCTTAACCTCTCGTCCGTCGTGGCAACCA 180
	Qy 181 GTCTCCCTCTTTATCGGACAGAGCGTTTCAGAGTCAACAGGGGCATTCGCCGTGAATT 240 Db 181 GTCTCCCTCTTTATCGGACAGAGCGTTTCAGAGTCAACAGGGGCATTCGCCGTGAATT 240
	Qy 241 CTCAAACTGGAGTCCAGAGTGGGTTCTGCCGGAAGAGAGACCTCAACATCAAGAT 300 Db 241 CTCAAACTGGAGTCCAGAGTGGGTTCTGCCGGAAGAGAGACCTCAACATCAAGAT 300
	Qy 301 GGATGTGGAACCTACGATCGCGCTCAGTCAGTTATTGGTGGGTCACTCCAGCCCCCAG 360 Db 301 GGATGTGGAACCTACGATCGCGCTCAGTCAGTTATTGGTGGGTCACTCCAGCCCCCAG 360
	Qy 361 TCCTTGAGGTCGTTGTTATACGGCTGGTCTTAGCCAGAGCTGGAGGAGAGATTC 420 Db 361 TCCTTGAGGTCGTTGTTATACGGCTGGTCTTAGCCAGAGCTGGAGGAGAGATTC 420
	Qy 421 GTCAGAGGTTCTTCAGATCAAGAAACTGAAGCCAACTTAACGCTTACAAGTTTGT 480 Db 421 GTCAGAGGTTCTTCAGATCAAGAAACTGAAGCCAACTTAACGCTTACAAGTTTGT 480
	Qy 481 ATTCGTAGTCAGGTTAACGATTGTCATGTCGTTAAACAGGAGAGGTGGCGTTCG 540 Db 481 ATTCGTAGTCAGGTTAACGATTGTCATGTCGTTAAACAGGAGAGGTGGCGTTCG 540
	Qy 541 GGGTTTGGTTTATAGCTCTACGCCACCATTCGCTACCCATTTCGAGGTTGTTCTGTGAA 600 Db 541 GGGTTTGGTTTATAGCTCTACGCCACCATTCGCTACCCATTTCGAGGTTGTTCTGTGAA 600
	Qy 601 AGCTACTGGGACAGACTTCATCCAAAGACTATGCTATTATCTGAGAGAAATAAGAC 660



similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are annotated by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arrian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES	source
repeat_region	
gene	
mRNA	
CDS	
gene	
mRNA	
CDS	
repeat_region	
gene	
mRNA	
CDS	

```

6527. .6629,6714. .6853,6937. .7023,7122. .7190,7278. .7349,
7429. .7547,7619. .7732,7814. .7859,7954. .8034)
/gene="T18K17.3"
/note="contains non-consensus splice site"
/codon_start=1
/product="putative serine carboxypeptidase; 5659-8034"
/protein_id="AAG52135.1"
/db_xref="GI:12324326"
/translation="MANKYFSSVLKSLLLHLVFLSKQHVDSASIVKFLPGFEGPLP
FELEGTIGIEEEVOLFYYFIKSERNPKEPDLILLWLTGGPGGSSISGLLFENGPLP
MKLDVYNTLSLSTVTSWTKSMIFLDQPVGTGFSYRTQKSPSPSGEAKRIH
EFLQKWLGRHOEFSSNPFFVGGDSYGMVVPATVQEISKGNVECCNPPINLOQYVLGN
PLTDVAIDNSRIPFAHGMALISDELFSKTKCKGVTNVHPRNTQCLKFIIEFNK
TNRILQIILDPLEETPCDYIYRLLTYTWANDATVREALQINKESIGEWVRCYRT
IPYNDIKSSMPYHVNNSISGRSLIYSGDHDLEVPYLGTQAWIRSLNYSIIDDRW
MKIKIAGYRTYANKMFTATIKGGHTIEKPEASIMFQWINGOPL"
/complement(6098. .6147)
/repeat_region
/repeat_id="AAG52135.1"
/gene="T18K17.4"
/note="similar to serine carboxypeptidase I precursor
GB:P07519 [Hordeum vulgare], glucose acyltransferase
GB:AAD01263 [Solanum berthaultii]; contains Pfam profile:
PF00450 Serine carboxypeptidase"
/join(<8937. .9082,9167. .9312,9394. .9470,9553. .9675,
9780. .9882,9952. .10091,10176. .10256,10362. .10430,
10527. .10598,10670. .10788,10877. .10990,11097. .11146,
11234. .>11310)
/gene="T18K17.4"
/join(8937. .9082,9167. .9312,9394. .9470,9553. .9675,
9780. .9882,9952. .10091,10176. .10256,10362. .10430,
10527. .10598,10670. .10788,10877. .10990,11097. .11146,
11234. .>11310)
/gene="T18K17.4"
/codon_start=1
/product="putative serine carboxypeptidase; 8937-11310"
/protein_id="AAG52136.1"
/db_xref="GI:12324327"
/translation="MANYISSVLKSLLLHLVFLSQHVDSASIVKFLPGFEGSLPF
ELETGTIGIEEEVOLFYYFIKSERNPKEPDLILLWLTGGPGGSSISGLLFENGPLM
MKLDVYNTLSLSTVTSWTKSMIFLDQPVGTGFSYRTQKSPSPSGEAKRIH
EFLQKWLGRHOEFSSNPFFVGGDSYGMVVPATVQEISKGNVECCNPPINLOQYVLGN
PLTDVAIDNSRIPFAHGMALISDELFSKTKCKGVTNVHPRNTQCLKFIIEFNK
TNRILQIILDPLEETPCDYIYRLLTYTWANDATVREALQINKESIGEWVRCYRT
IPYNDIKSSMPYHVNNSISGRSLIYSGDHDLEVPYLGTQAWIRSLNYSIIDDRW
MKIKIAGYRTYANKMFTATIKGGHTIEKPEASIMFQWINGOPL"
/complement(10483. .10509)
/repeat_region
/repeat_id="AAG52136.1"
/gene="T18K17.5"
/note="similar to serine carboxypeptidase I precursor
GB:P07519 [Hordeum vulgare], glucose acyltransferase
GB:AAD01263 [Solanum berthaultii]; contains Pfam profile:
PF00450 Serine carboxypeptidase"
/join(<12385. .12533,12631. .12776,12858. .12934,12999. .13121,
13208. .13310,13426. .13565,13643. .13729,13820. .13888,
13969. .14040,14131. .14249,14327. .14440,14520. .14565,
14657. .>14737)
/gene="T18K17.5"
/join(12385. .12533,12631. .12776,12858. .12934,12999. .13121,
13208. .13310,13426. .13565,13643. .13729,13820. .13888,
13969. .14040,14131. .14249,14327. .14440,14520. .14565,
14657. .>14737)
/gene="T18K17.5"
/codon_start=1
/product="putative serine carboxypeptidase; 12385-14737"
/protein_id="AAG52138.1"

```

```

/db_xref="GI:12324329"
/translation="MASNVFVSVLRSLILLIHTVFLGQHHVSSATIIKSLPGFEGPLP
FELETGTIGIEEEVOLFYYFIKSERNPKEPDLILLWLTGGPGGSSISGLLFENGPLA
MKLDVYNTLSLSTVTSWTKSMIFLDQPVGTGFSYRTQKSPSPSGEAKRIH
EFLQKWLGRHOEFSSNPFFVGGDSYGMVVPATVQEISKGNVECCNPPINLOQYVLGN
PLTDVYDYNRIPFAHGMALISDELFSKTKCKGVTNVHPRNTQCLKFIIEFNK
TNSICORRIIDPEFETETPCNYIYRLLAAWANDETVKAQIJKETIGEWVRCHYG
IPYNDIKSSIPYHVNNSINGYRSIYSGDHFVFLGTQAWIRSLNYSVIDDRWP
Query Match 29.5%; Score 239; DB 8; Length 70836;
Best Local Similarity 65.5%; Pred. No. 1.1e-51;
Matches 417; Conservative 0; Mismatches 205; Indels 15; Gaps 4;
QY 2 ATGAATCCTATGTTTACTTCTTCTTCCCTTTTACCACCTGTTTGGCCGGCAGCGCAAC 61
Db 19671 ATGAATCCTAAGTTTACTTGGTCTTAAACCGGCTTCTGCGCTCAAAAGCATAT 19730
QY 62 GCTGGACAGTTCGACACTGATGGTGATATATATTCACGCGCAGTTACTAGTTCTC 121
Db 19731 GGTG--CGGTGTGACATCGATGGAAACGCCATGTTCCAGNAAGTTACTAGTTCTC 19787
QY 122 CCCCTCATCTGCGGCCCTCAGAGTGGCGGCTAACTCTGCTCTCCCTCGTGGCAACAG 181
Db 19788 CCTGTCTATCCGTGGCC--GAGGCGGAGGCGTACTCTAGCAGGCGCGGTGGCGCA 19844
QY 182 TGTCCCTCTTTATCGACAGGACGTTTCAGAGTCAACAGGCGCATTCCTCCGTGAATTC 241
Db 19845 TGTCTTACGATATCTGTCAGGAATCTTCAGAAAGTTGATGAGGCGCATTCCTCGT 19904
QY 242 TCAAACTGGAGTCCAGAGTGGGTTGCTGCTCCCGCAAGAACCTCAACATCAAGATG 301
Db 19905 TCAAACTGGAGGCTTAAGTTGCGTTTCCGTCAGAACCTCAACATCAAGATG 19964
QY 302 GATGTGCAACCTAGGATCTCGGCTCAGTCTAGTCTATTTGGTGGTCACTCCAGCCCCAG 361
Db 19965 GACGTCCGAGCCAGATCTGCTCAACCTACTTGGCGGGTGGTGGTGGTGGTGGTGG 20024
QY 362 CCCTGGAGGTGTTGTTATGATGCGGTGCTTAAGCCAGCAAGCTGGAGAGAGATTCG 421
Db 20025 GAGAGGAGCAGTACTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 20084
QY 422 TCGAGGAGTTTCTTCCAGATCAAGAAACTGAAGCCAAACTTAACGCTTACAAAGTTTGA 481
Db 20085 TTGAAGAGTTTCTTCAAGATCGAGAAATCTGGAG-----AGGATGCTTACAAAGTTTGTG 20138
QY 482 TTCTGTAGTGGGTACGATGTCATGTCGTTGCGTGAAGAACCGAGGAAG--TGGCGTT 538
Db 20139 TTCTGTCTCGGACTTGGGACTCTGCAATCCAAATTCAGAGCGATGTCGGGATATTCATA 20198
QY 539 CGGGTTTGGTTTGTAGGCTCTAGCCACCATTTCGCTACCCCATTCGAGGTTGTGTTGTTG 598
Db 20199 GATGAATCTGGCGTTGCTGCTGTTTGGCTTTAAGCGATAAGCCGTTCTTGGTTATGTTCAA 20258
QY 599 AAAGTCTAGTGGGACAGACACTTCATCCAAAGACTATGT 635
Db 20259 AAAGTCTAATGTACCGAAGTTTCTGTTCCAAAGACTATGT 20295

```

## RESULT 5

AC116964

LOCUS

DEFINITION

AC116964

AC116964.1

VERSION

HTG: HTGS\_PHASE2.

KEYWORDS

Dictyostelium discoideum.

SOURCE

Dictyostelium discoideum

ORGANISM

Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

REFERENCE

1 (bases 1 to 256774)

AUTHORS

Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,

Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,

Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and

AC116964 256774 bp DNA linear HTG 04-APR-2002  
 Dictyostelium discoideum chromosome 2 map 3169953-3426725 strain  
 AX4, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

AC116964.1 GI:19920063

HTG: HTGS\_PHASE2.

Dictyostelium discoideum.

Dictyostelium discoideum

Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

1 (bases 1 to 256774)

Gloeckner, G., Eichinger, L., Szafranski, K., Pachebat, J., Dear, P.,

Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,

Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and







```

FNKQIIQKQEKIKNNKNEEVLKSVETIISPYLTGLMKYGLIIIEFLMWIISMIQ
YIIYKMMDDGSDGVENWPIKYYGLFIETIVNGWSAMKGYKVDQKDLFAAHL
LTMEFFYFFTYFYFYLYVYMLRGKRIKDEGMDKQWIAFYVSYAYVOTL
VIVTQIQPNOSGTDIIIGIAHDSLITCKENFKSGKRLNGHSKWLIPWLELNL
AQVLSLDVSTLACTLIGCHSEFNENFEYRCILIPLVIEFLIKPSFSFEPYCY
KFSNISSPSCHTAIASIVISISLSFQFGFLVMSMAAYRVYTRAHTVEDICAGFA
IGQVUNSTIYCFSEFFGLN"
Join(15996..16044,16634..18174)
/Note="ORF_ID:dd_00315"
/codon_start=1
/product="PUTATIVE CYTOSOLIC NADP-MALIC ENZYME"
/protein_id="AAL92636.1"
/db_xref="GI:19683013"
/translation="MNKPSILRNPSANKGLLPKPKVESLQEQSDRALSOFTSFTNL
ERYIFLNLORNETLFYLLSNNLELMPPIIYPTVGEACQKFGNEFRAQGYFAS
QDKNIRAMDNPAEGVDIIIVSDGSRILGLDGLTNGMGPVKGKQLYVAGAGFCP
TRTLVITDSNTKTKYLEDKXYLIGEBHPRIPOSEYPLVDEFLAARFKWPKVIVQF
EDISNDHCENLDEVRNKLFCNDIOQTSVILSGINAVRSQKRIKHRMVFILGA
GSAGIGVADCIMSLFDEAGVSKSEARKSFVWDSKGLITRDELTQKQKQYAREDY
TYQLKSLLEVYRDKPTAIIIGLSIGGSFSQEVIEEMAKHVEKPIVFAISNPTNAEC
TAEQAYQWTDGRCIFASGPEKPEYKGTKEVPGQGNMYPFGLGLAASVCEAKHYT
DAMIITAAKTLASFVEDSEVLTKIYPCGLQHIRIESTRIAVKVIKAYEEGMAQLPRP
DNIEALVKSQYVPSYDKKN"
Join(19036..19152,19298..19672,19768..20136)
/Note="ORF_ID:dd_00313"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92637.1"
/db_xref="GI:19683014"
/translation="MNIDKARVKLKSQSCNSFGSFESTYIPANLHGTMDDFEKTIL
KNINAKCNVLSKYLIVFLSLVGLLIVVGTFSANSVNSAGGFRMDAPQKNLND
FSIGSFSDFNNNNNDDDDSSPFGVGFVIALGFMFGCCIFGTSFAT
FRNKVLNIRDELAVANKITWNLESEVVRKYIPRHEYNGHKNQOYRNSIKY
GHEQRYREEVVHFIEIVPAKQIIFVPHVFSVMPKIOMNGTTSNNDGSLGPFSSNGG
NISIPMNSNF"
Join(20613..20705,21176..21646)
/Note="ORF_ID:dd_00312"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92638.1"
/db_xref="GI:19683015"
/translation="MNISRVNSTPKGSTFELTYTHLNLQIQSEVLNKLKIELDLIN
ENYKSNLKWLRDNLNFINDEELKKGKGTSMHGIKYDNNKRPYREIYFIEI
ESPFRQLNLTJKTFSSNQNHCDTIIIFQNNNSDNNNSDNNNNNNNNNNNNNN
NNNNNNKNIQDEKIGICINVDLKNF"
22386..25766
/Note="ORF_ID:dd_00310"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92639.1"
/db_xref="GI:19683016"
/translation="MNNDKPKPLTMDDDGDKTVAPPSSVIGGVSSRRPTSLRSQS
GINSSVDVAKTTTSSSSSTRSSSLSTSTATPKSTIIPSSSSSSSSSS
SSSSSSSSSTOKRPVLRKQSLVNGSAKPKSSSSSSSSSSSSSTSKAKTSG
DKQNTNTDVELKSNNAASFKEFDEPVSEFVRVSDFIILDEDDQSTPPPIID
SGNNVTTATTEDVDALYKQFLRQOQSIQDLIRQLEQIIQOQEQOQOQDQ
DQOQOQEQOQOQIIEQOGLSTFEMKQQAIDYERQOQNIIEKPKQEQQE
DEEVETIYVQPPITVELKTSWMSFFTKVLPSTNTSSSSSSSSSSSSSSSS
SSSSSSSSSSSSSSSSSSSNIAWTASTTTTISAATINNGVGLITTTTAT
ISNLBSSTSSSSSFTQINSSTLSTSPSLSTSSANISLPLSSSGIGYISVGGN
NNNNNNNNNNNNNNNNYDDDTNTIINNKGKRLRPPQIVHSQFOPHH
LQOQOQOQOQSFSTFSQOQIVPNIGYNSNNNNNNNNMMMLPHNQAGG
LALPFTITNNATIFOKYKSLWAFPTDFEKELETLMKQOYFIYFASVIVHVS
HSGKVGFLKHSPPSSSSQSTSPMLNQNPNNSNNNNNNVVVQDQSKVMVEEE
TEWSTIGOTNRYSOHTNDILVSPDRDOKLQHPNPHHHHHLPSVEYIKELNOR
LESSPFFHQHQLLPITDQFFQVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
LHMDNQINSGERKPVADRNKELAEIKKNEAYGNDKSLKLDYQHT
VAEYTSDFENIKRTIYIVPMYKLTTHYNNRDYFVINGQKSHGIRPYGASIKN
LAIKFTGGSVSSKVPNTLTKNLKGLVGLDNLQODIYNEDEQVQVLYNNE
LYLLPSSDQFLLYQMYIVLKNNAATTGIVLESKHSMEHIGGQVTLTPGTQTQF
SYKGTWCTEIKSGHNLQVIRIETNGGSDKPSGDLN"
complement(25915..26583)
/Note="ORF_ID:dd_00323"

```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92640.1"
/db_xref="GI:19683017"
/translation="MKFLITLLVVFILNLFVKSQVYVNFQPSNGKSGNSQIGIGY
SIQDCTCTIDRNNITMTSSNQWAKYSVENIPEVCSGATTTQTVKIGNCIKN
TNAFDESSLPAVQYFYKVISNQPLINSYIKSPMYNVDSCNENEVNIQYFSNAT
KAIEKTSRSPSSSYCSASNPYTYGCSINGCGPSYTSANCOQINSTPTLINENNYIS
VFCN"
27087..27299
/Note="ORF_ID:dd_00321"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92641.1"
/db_xref="GI:19683018"
/translation="MKDYETVTFSSIIFAFLFLRLCKFCVCCALCNVDPNVYGR
KRPGRSIVVEENEDDNGKEGLLVN"
complement(Join(27700..27770,28839..29324,29518..29584))
/Note="ORF_ID:dd_00319"

Query Match      6.5%  Score 52.2;  DB 2;  Length 59004;
Best Local Similarity 56.8%  Pred. No. 0.016;
Matches 96;  Conservative 0;  Mismatches 73;  Indels 0;  Gaps 0;

Qy 641 ATCTGAGAGAAATTAAGACACCTTAATAAGAGGATAAGTGTATTAACCTTACCTCTTAAT 700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55838 ATTTTAATTAATAAAAAAAAAAAAAAAAAAAAAATTAATAATTAATAATTAATAATAA 55897

Qy 701 AATAAACTCTATCTATGATGATGTTTCTTCTTCTCATCATCATCATCATGATGAA 760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55998 AAAAAAATGACCCACCAATGCTGTTTATTTTATTTTATTTTATTTTATTTTATTTT 55957

Qy 761 TAAACATCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 809
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55958 TAAACCAATCTCTATATATTAATAAAAAAAAAAGAGAAAAAAAAAAAA 56006

RESULT 9
AC005139 256172 bp DNA linear HTG 01-APR-1999
LOCUS Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
DEFINITION *** 5 unordered pieces.
AC005139
VERSION AC005139.3 GI:4558581
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 256172)
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 256172)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:437170.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 126535: contig of 126535 bp in length
* 126536 126735: gap of unknown length
* 126736 158845: contig of 32110 bp in length
* 158846 159045: gap of unknown length
* 159046 211538: contig of 52493 bp in length
* 211539 211738: gap of unknown length

```

*	211739		248106:	contig of 36368 bp in length
*	248107		248306:	gap of unknown length
*	248307		256172:	contig of 7866 bp in length.
<b>FEATURES</b>				
	source		Location/Qualifiers	
		1..256172	/organism="Plasmodium falciparum"	
		/db_xref="taxon:5833"		
		/chromosome="12"		
<b>BASE COUNT</b>	100700	a	26900 c	25826 g 101937 t      809 others
<b>ORIGIN</b>				
			Query Match	
			Best Local Similarity	6.4%: Score 52; DB 2; Length 256172;
			Matches 108; Conservative	0; Mismatches 75; Indels 1; Gaps 1;
<b>QY</b>	626	AAGACTATGCTTATTATCTGCAGAGAAATTAAGACCACTTAATAAAGA/GCATAAGTGTTA	685	
<b>Db</b>	186868	AAGAAAAATTGAATATAAATAATAAATAAANAANAAAAGAAATA-TATATATCAT	186926	
<b>QY</b>	686	TAACTTAGCTCTAATATAATAAACCTCTATCTATGTATGTATGGTTTTTC/TTCATCGATCA	745	
<b>Db</b>	186927	ATTTCTCATGTAACC	186986	
<b>QY</b>	746	TCATCATGTATGGATAAACA/CCTTCCCTTGTCTTAAANAAAAAAAAAAAAAAAAA	805	
<b>Db</b>	186987	GTTTATTATGTGTTAAAAATTATATATTTATTTTACAAAAAANAANAAAATAA	187046	
<b>QY</b>	806	AAAA	809	
<b>Db</b>	187047	TAAA	187050	
<b>RESULT 10</b>	CNS08CB9			
<b>LOCUS</b>	Oryza sativa chromosome 12 clone OSJNBb010I110. *** SEQUENCING IN PROGRESS ***, 6 ordered pieces.			
<b>DEFINITION</b>	AL845347			
<b>ACCESSION</b>	AL845347.1 GI:22208188			
<b>VERSION</b>	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.			
<b>KEYWORDS</b>	Oryza sativa.			
<b>SOURCE</b>	Oryza sativa			
<b>ORGANISM</b>	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryziales; Oryza.			
<b>REFERENCE</b>	1 (bases 1 to 131665)			
<b>AUTHORS</b>	Choisne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P., Seguren,B., Pelletier,E., Scarpelli,C., Salanoubat,M., Weissenbach,J. and Quetier,F.			
	Oryza sativa chromosome 12 sequencing Unpublished			
	2 (bases 1 to 131665)			
<b>TITLE</b>	Genoscope.			
<b>JOURNAL</b>	Direct Submission			
	BP 1911 91006 ERY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
<b>COMMENT</b>	IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. Contigs composition : 16662 bp contig from 1 to 16662 6571 bp contig from 16763 to 23333 56308 bp contig from 23434 to 79741 32083 bp contig from 79842 to 111924 8612 bp contig from 112025 to 120636 10929 bp contig from 120737 to 131665.  * NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces			



```
EEERKVKAWDKLHLHLEHTPSSKYLTSILKSRADKSIQIKDKDNKNTISDKENIAK
REVEFYQOEKEDNEETHKLLKEKVEDVDLKKLEIDRPIRIHEVTRAKTSSIH
KSPGLDGNALFYKHYHNSIARILTIAFNDLTLTKKEIPTKFKEGVITIPKGDLEIN
ISNRKPNTLLNTDYDKILSKILNSRLDITSKILNKFGQFVFNRFIOQDIOIMEVIE
ISNRKNTLLITFYDNKAFDSISKHSITRLEHLGIPPKFTAILLNLKDKTKKIKI
NDELVNGITIRGTGQKQDIPSTPTIFALVLEPLLDIDINDNTIGFTIPNSKSLTAF
ADJIAFTNSTEELMKINTKIOKVCYSATSSLSNKEKTVMAIGADKPHDLPPQESTVPE
RYLGUNFTKTLGNSKYNLTIOEMKNLKKWSQATIMKAKMTILKTYVLSKLTTHQYM
DNLNEQIEENNTWRFVFSVKNYTYEERKYTKMTKIDRAYADWKEGGKLDWIDEL
RHIAFKIWMNRLLHNNNNNNNTLQEWYMEQLSRKKPTLPQSPICADGTGVYSESNT
KTIOH"
Join(18321..18339,18404..18924)
/Note="ORF_ID:dd_00446"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA09331.1"
/db_xref="GI:20066217"
/translation="MAEYVGRNLTFRFFSRSLPGINHERTRCKICGHIFRDPYSHL
FLCDIIDLKTIISTVKNLUSFKIQHWSMDTLDISKYNTERIFPNLIGIIAHQLW
KLICKLFDTEDESKPEPEKQVETELLNLIETKEFTTLKKIKHDEALLKNTQDLH
RYKENKAWQTPAAPNPLPI"
1886..20788
/Note="ORF_ID:dd_00444"
/codon_start=1
/product="Putative Nuclear protein"
/protein_id="AA09332.1"
/db_xref="GI:20066218"
/translation="MNNFLIISIIIVSVGECALDFSSISVIGISDSQWSCIA
SNORVITQVMSGGOYNSIVSVYARAEAGPNIDLYAFCLSECQGNTPASSAIGS
LYVSLKSGINFMWIDVQECDCGWCAGSNADYVQEAETAGGLVGVYVSSEGE
WPQTVGNLUSTLSQPLWYAHYDDNPSDFSTAFYEFQWTSIPAKQYIGNTNQCQSVSD
LDFYGGSGSTSGSGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGSAGS
SGSGSGSGSGSGSAGSGTGGSGSI"
complement(join(21433..23231,23387..23705))
/Note="ORF_ID:dd_00442"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA09333.1"
/db_xref="GI:20066219"
/translation="MSINKEKOKIINEINPEKPNISGDDNNIILIRETGNEDCTD
IPTPTPIIDVELISPLSPPPPTTIIDNNKTPIITPOLIDFDLPSSATLIHSFP
TNSRTILGSHVHIKSGSDEYSEIIRKWKPTVFVELCSSRAGILFNTNDDSDIYYN
NNNNNNKNNVSNKSKRTKILDRGTSVSRRLPQREKHFLDLDLDDQDFSSASTSK
FVLETEIDIDEDMDQDNELYTSEETEODENFDODDYGYEDYLDHNNINHLMN
QPRNFYLYDDNDLIEBNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
KNKSNLTKTLKINNFKNIFNSDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
FINNLNNSFNTSSDADDERSHSPESITFTQYIDNEEEEDNEEENNEENNE
EIKNDQSEISIKDIIRIKCNGLSGVLHILIAKLKAKKSKVGPGEFTAFLEA
RKISLVYLBQRQGITLQRVNSLSWLEKIKFYFLFMASLSLSEITTEEDAIKNSD
ELVKNLNEFRGFPSPVHTITVTERDQYMAARLMCFCKKIVAVYVKGHIGGIIREWE
NNNINLEFESVNIKKSNLNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
WASLILPIGILGYFFNKKFNFI"
complement(25734..26420)
/Note="ORF_ID:dd_01644"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA09335.1"
/db_xref="GI:20066221"
/translation="MFEINKIYMNASERDKIITSKIOKIDKEILPVCASPNVYN
DIDFONKLNQNLNKKKIKNOGIVSDISLPIKKNELIQNKLSLSENGNLKNE
KTESFINDKLIVDKDKNLGVGRYDVHYCVFGYQLKKSIVIRKPDNPPIISSDL
LAITYKAYQYMKQKQWFDYKKTRINISNPSPIGYRIGINDIHYNGLSTIIEFDTEL
VCSFDDVPDK"
complement(27154..27738)
/Note="ORF_ID:dd_01643"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA09336.1"
/db_xref="GI:20066222"
/translation="MENYLFRDANSEELKETEGLKPKIKLENILKIDGAPNKQITLLE
VYKNYKDALSKSGLESFNOPNOTDOLFKEVYLIEGGDNTKHYCYCIDAYTANP
TVIKVTRPDANSLOPLTNVELLAIYSKAFQFIYEEEEKVGNPGHIOGMLNRGVSHOR
YGIWGHLDLUVYNGFSDVLIYKDFIFCEFRVDS"
```

```
CDS
complement(join(28103..28128,28441..28881,28964..29004,
29963..30505,30619..30989))
/Note="ORF_ID:dd_01642"
/codon_start=1
/product="Putative GTPase"
/protein_id="AA09337.1"
/db_xref="GI:20066223"
/translation="MKYINLIFLICYINQOSIGADRNEFFPINEILAAKSLKSLYN
FEAKTLEECFYCGFCVSPYEPKTYGITTIRNSTRLSFVDSFISFRLSELIIIEE
NIOIQIVFYKTLPLLKNLELLEVSKQEPFDFNSPPKLTQVQNSISVPLUSIWF
EGVINRIIVVKTLPFGKYPKLTKINTYINTLSLNHDINTVPSMDLFPNLELINFQI
HNEMSOKGYKFSIDISINQWYKRVNMAEQFINSGEATIQKFTQSLISRFKLDY
LLIDGFTVDPDSIGYVNFSDKDKGFLHLFKFVLGAGSGVGKTKILFRITDDPTE
THITIGIEFKITVYIEGPKIQIWDTAGERFRVHNSHYRGCHGVIIIVDYVDORS
FENVPSWIEDIRRYANENVIKIIIGNKNDLVSKVVDVPLAQEFADSLDIPKFEISAK
QSNIEEAFISLVKLFKLNNT"
complement(32023..32289)
/Note="ORF_ID:dd_01641"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA09338.1"
/db_xref="GI:20066224"
/translation="MKCLKKVRIFOLIOIEYINVALEFNFINSQNEATIQKFTSQ
SLISKFKLWLVFKIGISFTLDPSIGYLFNNSKMDIYIGIDLIV"
complement(join(34992..35316,35500..35717))
Query Match 6.3%; Score 50.8; DB 2; Length 156533;
Best Local Similarity 57.6%; Pred. No. 0.044;
Matches 91; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 651 AATTAAAGACCACTTAATAAGAGGATAAGTGTATTAACCTCACTTAATAATAAACTC 710
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92522 AATTCAAGACAATGATATTAATAATAATAATAATAATAATAATAATAATAATTC 92463
QY 711 TATCATGTATGATGTTTTCTTTCTTCATCATCATCATCATCATCATCATCATCAT 770
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92462 ATTTTATTATTAATATTTCTATTTTAAATAATAATAATAATAATAATAATAATA 92403
QY 771 TTCCTTTGTTCTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 808
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 92402 TTATCTATTGATAGATAAAAAAAAAAAAAAAAAAAAAA 92365

RESULT 13
DDU67089/c DDU67089 1618 bp DNA linear INV 26-SEP-1996
LOCUS DDU67089 Dictyostelium discoideum unknown protein (DG1008) gene, partial
DEFINITION Dictyostelium discoideum.
ACCESSION U67089
VERSION U67089
KEYWORDS Dictyostelium discoideum.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 1618)
AUTHORS Loomis,W.F.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-1996) Dpt. of Biology 0322, University of
California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-032,
USA
FEATURES
source Location/Qualifiers
1..1618
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
265..1618
/gene="DG1008"
265..1618
/gene="DG1008"
/codon_start=1
/product="unknown"
/protein_id="AA08873.1"
/db_xref="GI:1561772"
gene
CDS
```



Db 80018 TTCTTCATATCCTATTCTCTATTTGTTATTCACACATGTTTCATGTTAAATGA 80077  
QY 679 AGTGTATTAACCTTACCTCTAATAATAAACTCTATCTATGTATGTATGTTCTTTGTTC 738  
Db 80078 ATATAAAAAATATATGTATTTAAATATATATATATATATATATATATATATAT 80137  
QY 739 TCGATCATCATCATGTATGGAATAAAACATCTTTCTTTGTTCTTAAATAAAAAA 798  
Db 80138 CAAAAAGGTTGTATATATGAAAAAATATATTTCTTTTATACAAATGAAAAA 80197  
QY 799 AAAAAA 806  
Db 80198 AAAAAA 80205

Search completed: January 25, 2003, 12:56:34  
Job time : 4443 secs

